

A;Introns: 183/1; 247/1; 288/1; 318/1; 364/1; 388/2; 457/1; 481/2; 521/3

Query Match 30.2%; Score 118; DB 2; Length 561;
Best Local Similarity 36.8%; Pred. No. 0.00025;
Matches 21; Conservative 11; Mismatches 23; Indels 2; Gaps 2;

QY 12 CPSNEIFSRCDGRC-QRQPCNV-VPKPLCIKICAPGCVCRLGYLRNKKVCPVRSKC 66
 ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 123 CPVNEVSNECHNPCTERKKCPQKNAPQVNCLMACQVGCGMDFGRNNQGVCKVEAEC 179

RESULT 3

T27319
hypothetical protein Y69H2.3b - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T27319
R;McMurray, A.
submitted to the EMBL Data Library, August 1997
A;Reference number: Z20343
A;Accession: T27319
A;Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-626 <WIL>
A;Cross-references: EMBL:X98877; PIDN:CAB54473.1; GSPDB:GN00023; CESP:Y69H2.3b
A;Experimental source: clone Y69H2
C;Genetics:
A;Gene: CESP:Y69H2.3b
A;Map position: 5
A;Introns: 183/1; 247/1; 312/1; 353/1; 383/1; 429/1; 453/2; 522/1; 546/2; 586/3

Query Match 30.2%; Score 118; DB 2; Length 626;
Best Local Similarity 36.8%; Pred. No. 0.00027;
Matches 21; Conservative 11; Mismatches 23; Indels 2; Gaps 2;

QY 12 CPSNEIFSRCDGRC-QRQPCNV-VPKPLCIKICAPGCVCRLGYLRNKKVCPVRSKC 66
 ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 123 CPVNEVSNECHNPCTERKKCPQKNAPQVNCLMACQVGCGMDFGRNNQGVCKVEAEC 179

RESULT 4

T30197
alpha tectorin - mouse
C;Species: Mus musculus (house mouse)
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jul-2000
C;Accession: T30197
R;Legan, P.K.; Rau, A.; Keene, J.N.; Richardson, G.P.
J. Biol. Chem. 272, 8791-8801, 1997
A;Title: The mouse tectorins. Modular matrix proteins of the inner ear homologous
A;Reference number: Z20771; MUID:97236843; PMID:9079715
A;Accession: T30197
A;Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2155 <LEG>
A;Cross-references: EMBL:X99805; NID:gl915908; PIDN:CAA68138.1; PID:gl915909
A;Experimental source: strain CD1; whole cochlea
A;Note: non-collagenous protein only expressed in the inner ear, by cells both

Query Match 28.9%; Score 113; DB 2; Length 2155;
Best Local Similarity 38.6%; Pred. No. 0.0021;
Matches 22; Conservative 7; Mismatches 26; Indels 2; Gaps 2;

QY 11 KCPNSIEFSRDGRCQRQPCNVVPKPLGIKICAPGCVCRLGYLRNKKVCPVRSKC 67
 ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 983 ECPENSHEEC-MTCTETCEIALGPICVDSCSECCDEGY-RLOGSQCVTRSEC 1037

RESULT 5

T42215
zonadhesin - mouse
N;Alternate names: sperm-specific membrane protein
C;Species: Mus musculus (house mouse)
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text change 03-Dec-1999

Matches	22;	Conservative	9;	Mismatches	18;	Indels	17;	Gaps	3;
Qy	11	KPSNEIFRCDGRC-----QRFCNVVPKPLCIKICAPGCVCRGLGYLRNKKVCV	61						
Db	1455	KCPGSGSYSTCANPCPCATCLSLNNPSYCPSTLP-----CAEGCECQKHILSGTS-CV	1506						
Qy	62	PRSKCG	67						
Db	1507	PLSQCG	1512						
RESULT 10									
T15609									
		hypothetical protein C25E10.8 - Caenorhabditis elegans							
		C/Species: Caenorhabditis elegans							
		C/Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999							
		C/Accession: T15609							
		R:Bradshaw, H.							
		submitted to the EMBL Data Library, February 1996							
		A/Description: The sequence of C. elegans cosmid C25E10.							
		A/Reference number: Z18376							
		A/Accession: T15609							
		A/Status: preliminary; translated from GB/EMBL/DBJ							
		A/Molecule type: DNA							
		A/Residues: 1-137 <BRA>							
		A/Cross-references: EMBL:U50311; NID:g1226295; PTD:g1226303; PIDN:AAA92313.1; CESP:C25E10.8							
		A/Experimental source: strain Bristol N2							
		C/Genetics:							
		A/Gene: CESP:C25E10.8							
		A/Introns: 1/3; 19/1; 42/3							
Query Match	25.7%;	Score 100.5;	DB 2;	Length 137;					
Best Local Similarity	36.8%;	Pred. No. 0.0045;							
Matches	21;	Conservative	7;	Mismatches	26;	Indels	3;	Gaps	3;
Qy	11	KPSNEIFRCDGRCQRCQRCFNVVPKPLCIKICAPG-CVCRGLGYLRNKKVCVPRSKC	66						
Db	81	KCPENETFRCTACPTCEKQGRP-CTRQCIVVCQSSGFVRNGYR-CTELKRC	135						
RESULT 11									
T16574									
		hypothetical protein K05F1.10 - Caenorhabditis elegans							
		C/Species: Caenorhabditis elegans							
		C/Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999							
		C/Accession: T16574							
		R:Wohldmann, P.							
		submitted to the EMBL Data Library, June 1995							
		A/Description: The sequence of C. elegans cosmid K05F1.							
		A/Reference number: Z18537							
		A/Accession: T16574							
		A/Status: preliminary; translated from GB/EMBL/DBJ							
		A/Molecule type: DNA							
		A/Residues: 1-140 <WOH>							
		A/Cross-references: EMBL:U29377; NID:g858173; PID:g858180; PIDN:AAA68717.1; CESP:K05F1.10							
		A/Experimental source: strain Bristol N2							
		C/Genetics:							
		A/Gene: CESP:K05F1.10							
		A/Introns: 31/3; 77/2							
Query Match	25.6%;	Score 100;	DB 2;	Length 140;					
Best Local Similarity	32.8%;	Pred. No. 0.0051;							
Matches	21;	Conservative	10;	Mismatches	25;	Indels	8;	Gaps	2;
Qy	11	KPSNEIFRSC--DGRQRCFCNVVPKPLCI-----KICAPGCVCRGLGYLRNKKVCV	62						
Db	52	ECQKHEHLICGPERHCDRTCENLFPHPHCLNHLHAKYFPRPCVNDGTVRSEKGCIR	111						
Qy	63	RSKC	66						
Db	112	PSHC	115						

RESULT 12

VWHU
 von Willebrand factor precursor - human
 C/Species: Homo sapiens (man)
 C/Date: 04-Dec-1986 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
 C/Accession: A34480; S02377; A37139; S23676; A25298; A25366; S23618; S23645; A94
 R/Mancuso, D.J.; Tuley, E.A.; Westfield, L.A.; Worrall, N.K.; Shelton-Inloes, B.B.; Sora
 J. Biol. Chem. 264, 19514-19527, 1989
 A/Title: Structure of the gene for human von Willebrand factor.
 A/Reference number: A34480; MUID:90062044; PMID:2584182
 A/Accession: A34480
 A/Molecule type: DNA
 A/Residues: 1-2813 <MAN>
 A/Cross-references: EMBL:M25864
 R/Bonthron, D.; Orkin, S.H.
 Eur. J. Biochem. 171, 51-57, 1988
 A/Title: The human von Willebrand factor gene. Structure of the 5' region.
 A/Reference number: S02377; MUID:88111704; PMID:2828057
 A/Accession: S02377
 A/Molecule type: DNA
 A/Residues: 1-177 <BO2>
 A/Cross-references: EMBL:X06828
 R/Mancuso, D.J.; Tuley, E.A.; Westfield, L.A.; Lester-Mancuso, T.L.; Le Beau, M.M.; Sora
 Biochemistry 30, 253-269, 1991
 A/Title: Human von Willebrand factor gene and pseudogene: structural analysis and differ
 A/Reference number: A37139; MUID:91105089; PMID:1988024
 A/Accession: A37139
 A/Molecule type: DNA
 A/Residues: 990-1947 <MAD>
 A/Cross-references: GB:M0675; NID:g340357; PIDN:AAA61295.1; PID:g553810
 A/Note: the authors translated the Codon CGC for residue 156 as Gln
 R/Collins, C.J.; Underdahl, J.P.; Levene, R.B.; Ravera, C.P.; Morin, M.J.; Dombalagian,
 Proc. Natl. Acad. Sci. U.S.A. 84, 4393-4397, 1987
 A/Title: Molecular cloning of the human gene for von Willebrand factor and identificatio
 A/Reference number: S23676; MUID:87260814; PMID:3496594
 A/Accession: S23676
 A/Molecule type: DNA
 A/Residues: 2731-2813 <COL>
 A/Cross-references: EMBL:M16945
 R/Bonthron, D.; Orr, E.C.; Mitscock, L.M.; Ginsburg, D.; Handin, R.I.; Orkin, S.H.
 Nucleic Acids Res. 14, 7125-7127, 1986
 A/Title: Nucleotide sequence of pre-pro-von Willebrand factor cDNA.
 A/Reference number: A25298; MUID:87016349; PMID:3489923
 A/Accession: A25298
 A/Molecule type: mRNA
 A/Residues: 1-470,'V', 472-2813 <BON>
 A/Cross-references: EMBL:X04385
 R/Verweij, C.L.; Diergaarde, P.J.; Hart, M.; Pannekoek, H.
 EMBO J. 5, 1839-1847, 1986
 A/Title: Full-length von Willebrand factor (vWF) cDNA encodes a highly repetitive protei
 A/Reference number: A91044; MUID:870004550; PMID:3019665
 A/Accession: A25469
 A/Molecule type: mRNA
 A/Residues: 1-470,'V', 472-483,'R', 485-1022,'K', 1024-1025,'E', 1027-1400 <VER>
 A/Cross-references: EMBL:X04146
 A/Note: this sequence has been revised in reference A91056
 R/Verweij, C.L.; Diergaarde, P.J.; Hart, M.; Pannekoek, H.
 EMBO J. 5, 3074, 1986
 A/Reference number: A91056
 A/Accession: A25366
 A/Molecule type: mRNA
 A/Residues: 1021-1030 <VE2>
 A/Note: this is a revision to the sequence from reference A91044
 R/Shelton-Inloes, B.B.; Broze Jr., G.J.; Miletic, J.P.; Sadler, J.E.
 Biochem. Biophys. Res. Commun. 144, 657-665, 1987
 A/Title: Evolution of human von Willebrand factor: cDNA sequence polymorphisms, repeated
 A/Reference number: S23618; MUID:87213253; PMID:3495266
 A/Accession: S23618
 A/Molecule type: mRNA
 A/Residues: 1-120 <SH2>
 A/Cross-references: EMBL:M17588; NID:g799330; PIDN:AAA65940.1; PID:g340316
 A/Accession: S23645
 A/Molecule type: protein
 A/Cross-references: GDB:119125; OMIM:193400

A/Residues: 23-56 <SH3>
 R/Sadler, J.E.; Shelton-Inloes, B.B.; Sorace, J.M.; Harlan, J.M.; Titani, K.; Davie,
 Proc. Natl. Acad. Sci. U.S.A. 82, 6394-6398, 1985
 A/Title: Cloning and characterization of two cDNAs coding for human von Willebrand f.
 A/Reference number: A94060; MUID:86016708; PMID:2864688
 A/Accession: A94060
 A/Molecule type: mRNA
 A/Residues: 'WA', 739,'C', 744-769,'H', 771-788,'A', 790-803,'S', 805-873; 1289-1471,'D', 1-
 A/Note: the authors translated the codon TCG for residue 2168 as Cys
 R/Shelton-Inloes, B.B.; Titani, K.; Sadler, J.E.
 Biochemistry 25, 3164-3171, 1986
 A/Title: cDNA sequences for human von Willebrand factor reveal five types of repeat:
 A/Reference number: A90504; MUID:86269894; PMID:3488076
 A/Accession: A90504
 A/Molecule type: mRNA
 A/Residues: 781-788,'A', 790-1424 <SHE>
 A/Note: 852-Gln, 857-Asp, and 1381-Thr were also found
 R/Ginsburg, D.; Handin, R.I.; Bonthron, D.T.; Donlon, T.A.; Bruns, G.A.P.; Latt, S.A
 Science 228, 1401-1406, 1985
 A/Title: Human von Willebrand factor (vWF): isolation of complementary DNA (cDNA) cl.
 A/Reference number: A44178; MUID:85244583; PMID:3874428
 A/Accession: A44178
 A/Molecule type: mRNA
 A/Residues: 2621-2813 <GIN>
 A/Cross-references: EMBL:X03028; NID:g340308; PIDN:AAA61293.1; PID:g340309
 R/Verweij, C.L.; de Vries, C.J.M.; Distel, B.; van Zonneveld, A.J.; van Kessel, A.G.
 Nucleic Acids Res. 13, 4699-4717, 1985
 A/Title: Construction of cDNA coding for human von Willebrand factor using antibody
 A/Reference number: S07363; MUID:85269603; PMID:3875078
 A/Accession: S07363
 A/Molecule type: mRNA
 A/Residues: 2731-2813 <VE3>
 A/Cross-references: EMBL:X02672; NID:g37939; PIDN:CAA26503.1; PID:g37940
 R/Lynch, D.C.; Zimmermann, T.S.; Collins, C.J.; Brown, M.; Morin, M.J.; Ling, E.H.; L
 Cell 41, 49-56, 1985
 A/Title: Molecular cloning of cDNA for human von Willebrand factor: authentication b
 A/Reference number: S23678; MUID:85201687; PMID:3873280
 A/Accession: S23678
 A/Molecule type: mRNA
 A/Residues: 2731-2813 <LYN>
 A/Cross-references: EMBL:X03028
 R/Titani, K.; Kumar, S.; Takio, K.; Ericsson, L.H.; Wade, R.D.; Ashida, K.; Walsh, K
 Biochemistry 25, 3171-3184, 1986
 A/Title: Amino acid sequences of human von Willebrand factor.
 A/Reference number: A90505; MUID:86269895; PMID:3524673
 A/Accession: A90505
 A/Molecule type: protein
 A/Residues: 764-788,'A', 790-1471,'D', 1473-2813 <TIT>
 A/Note: 789-Thr was also found
 R/Chopek, M.W.; Girma, J.P.; Fujikawa, K.; Davie, E.W.; Titani, K.
 Biochemistry 25, 3146-3155, 1986
 A/Title: Human von Willebrand factor: a multivalent protein composed of identical su
 A/Reference number: A23464; MUID:86269892; PMID:3015199
 A/Accession: A23464
 A/Molecule type: protein
 A/Residues: 764-773; 2803-2813 <CHO>
 R/Dent, J.A.; Berkowitz, S.D.; Ware, J.; Kasper, C.K.; Ruggeri, Z.M.
 Proc. Natl. Acad. Sci. U.S.A. 87, 6308-6310, 1990
 A/Title: Identification of a cleavage site directing the immunochemical detection of
 A/Reference number: A36013; MUID:90349604; PMID:2385594
 A/Accession: A36013
 A/Molecule type: protein
 A/Residues: 1606-1617 <DEN>
 R/Fay, P.J.; Kawai, Y.; Wagner, D.D.; Ginsburg, D.; Bonthron, D.; Ohlsson-Wilhelm, B
 Science 232, 995-998, 1986
 A/Title: Propolypeptide of von Willebrand factor circulates in blood and is identica
 A/Reference number: A60913; MUID:86208144; PMID:3486471
 A/Accession: A60913
 A/Molecule type: protein
 A/Residues: 576-590 <FAY>
 C/Genetics:
 A/Gene: GDB:VWF
 A/Cross-references: GDB:119125; OMIM:193400

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2003, 08:09:22 ; Search time 29 seconds
(without alignments)
421.775 Million cell updates/sec

Title: US-09-506-978-1

Perfect score: 391

Sequence: 1 GGFGLGGKCPSEIFSR.....CRLGYLRNKKVCVPRSKG 67

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/prodata/2/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/prodata/2/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/prodata/2/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/prodata/2/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/prodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
- 7: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/prodata/2/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/prodata/2/pubpaa/US09A_PUBCOMB.pep:*
- 10: /cgn2_6/prodata/2/pubpaa/US09B_PUBCOMB.pep:*
- 11: /cgn2_6/prodata/2/pubpaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/prodata/2/pubpaa/US09_NEW_PUB.pep:*
- 13: /cgn2_6/prodata/2/pubpaa/US10A_PUBCOMB.pep:*
- 14: /cgn2_6/prodata/2/pubpaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/prodata/2/pubpaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/prodata/2/pubpaa/US10_NEW_PUB.pep:*
- 17: /cgn2_6/prodata/2/pubpaa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/prodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	391	100.0	67	12	US-10-174-151-1
2	391	100.0	69	12	US-10-174-151-2
3	391	100.0	71	12	US-10-174-151-3
4	391	100.0	73	12	US-10-174-151-4
5	119	30.4	67	12	US-10-087-887-40
6	106.5	27.2	5405	9	US-09-922-217-1116
7	106.5	27.2	5405	14	US-10-025-380-1116
8	100	25.6	84	11	US-09-498-272-59
9	100	25.6	91	11	US-09-498-272-128
10	98.5	25.2	76	12	US-10-029-386-31899
11	98	25.1	115	12	US-10-114-774-4
12	98	25.1	2813	10	US-09-381-261A-1
13	98	25.1	2813	12	US-10-021-660-79
14	98	25.1	2813	15	US-10-020-141-4
15	91	23.3	82	11	US-09-498-272-25

16	91	23.3	82	11	US-09-498-272-49	Sequence 49, Appl
17	91	23.3	171	11	US-09-498-272-62	Sequence 62, Appl
18	90	23.0	4123	15	US-10-213-509-5	Sequence 5, Appl
19	88.5	22.6	72	12	US-10-029-386-32016	Sequence 32016, A
20	87.5	22.4	2813	10	US-09-886-900-2	Sequence 2, Appl
21	87	22.3	75	11	US-09-498-272-57	Sequence 57, Appl
22	87	22.3	94	11	US-09-498-272-22	Sequence 22, Appl
23	85	21.7	759	15	US-10-189-971-22	Sequence 22, Appl
24	85	21.7	1057	15	US-10-189-971-6	Sequence 6, Appl
25	85	21.7	1192	15	US-10-189-971-18	Sequence 18, Appl
26	85	21.7	1207	15	US-10-189-971-20	Sequence 20, Appl
27	85	21.7	1251	15	US-10-189-971-16	Sequence 16, Appl
28	85	21.7	1342	15	US-10-189-971-24	Sequence 24, Appl
29	85	21.7	1477	15	US-10-189-971-8	Sequence 8, Appl
30	85	21.7	1512	15	US-10-189-971-10	Sequence 10, Appl
31	85	21.7	1535	15	US-10-189-971-14	Sequence 14, Appl
32	85	21.7	1570	15	US-10-189-971-12	Sequence 12, Appl
33	85	21.7	1593	15	US-10-189-971-4	Sequence 4, Appl
34	85	21.7	1628	15	US-10-189-971-2	Sequence 2, Appl
35	83.5	21.4	355	9	US-09-826-212-14	Sequence 14, Appl
36	83.5	21.4	355	10	US-09-935-727-16	Sequence 16, Appl
37	83.5	21.4	355	15	US-10-186-643-14	Sequence 14, Appl
38	82.5	21.1	469	9	US-09-925-301-1279	Sequence 1279, Ap
39	82.5	21.1	469	15	US-10-106-698-6377	Sequence 6377, Ap
40	82.5	21.1	482	15	US-10-106-698-4627	Sequence 4627, Ap
41	82	21.0	61	11	US-09-498-272-61	Sequence 61, Appl
42	81	20.7	2914	12	US-10-093-463-82	Sequence 82, Appl
43	80	20.5	74	11	US-09-498-272-42	Sequence 42, Appl
44	80	20.5	350	15	US-10-046-433-41	Sequence 41, Appl
45	78.5	20.1	77	10	US-09-950-933A-46	Sequence 46, Appl

ALIGNMENTS

RESULT 1

US-10-174-151-1
; Sequence 1, Application US/10174151
; Publication No. US20030165514A1
; GENERAL INFORMATION:
; APPLICANT: Spertini, Francois
; TITLE OF INVENTION: NOVEL BEE VENOM POLYPEPTIDES AND METHODS OF USE THEREOF
; FILE REFERENCE: 18519-001
; CURRENT APPLICATION NUMBER: US/10/174,151
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US/09/506,978
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 67
; TYPE: PRT
; ORGANISM: Apis mellifera
US-10-174-151-1

Query Match 100.0%; Score 391; DB 12; Length 67;
Best Local Similarity 100.0%; Pred. No. 2.3e-34;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	GGFGGLGGKCPSEIFSRCDGRCQFCNPVVVKPLCIKICAPGCVRLGYLRNKKVC	60
Db	1	GGFGGLGGKCPSEIFSRCDGRCQFCNPVVVKPLCIKICAPGCVRLGYLRNKKVC	60
Qy	61	VPRSKG 67	
Db	61	VPRSKG 67	

RESULT 2

US-10-174-151-2
; Sequence 2, Application US/10174151
; Publication No. US20030165514A1
; GENERAL INFORMATION:


```
; APPLICANT: Lung, Oliver
; APPLICANT: Tram, Khanh-Uyen
; TITLE OF INVENTION: GENES FOR MALE ACCESSORY GLAND PROTEINS IN DROSOPHILA
; TITLE OF INVENTION: MELANOGASTER
; FILE REFERENCE: 19603/1791
; CURRENT APPLICATION NUMBER: US/10/114,774
; CURRENT FILING DATE: 2002-04-02
; PRIOR APPLICATION NUMBER: 09/219,983
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 4
; LENGTH: 115
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-114-774-4

Query Match      25.1%; Score 98; DB 12; Length 115;
Best Local Similarity 34.4%; Pred. No. 0.0039;
Matches 22; Conservative 9; Mismatches 31; Indels 2; Gaps 2;

QY 4 GGLGGKCPSPNEIFSRCDGRQCFPCNVVPKPLCIKICAPGVCRLGYLRKK-KVCVP 62
Db 26 GWQPKVDTANGTQTECPVACPETCEYSNGNP-CVKGAPCVCKRPGYVNERIPACVL 84

QY 63 RSKC 66
Db 85 RSDC 88

RESULT 12
US-09-381-261A-1
; Sequence 1, Application US/09381261A
; Patent No. US20020123457A1
; GENERAL INFORMATION:
; APPLICANT: Loscalzo, Joseph
; APPLICANT: Inbal, Aida
; TITLE OF INVENTION: No. US20020123457A1e1 Anti-Platelet Agent
; FILE REFERENCE: 102258.327
; CURRENT APPLICATION NUMBER: US/09/381,261A
; CURRENT FILING DATE: 1999-12-07
; PRIOR APPLICATION NUMBER: PCT/US98/06092
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/046,981
; PRIOR FILING DATE: 1997-03-27
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 1
; LENGTH: 2813
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-381-261A-1

Query Match      25.1%; Score 98; DB 10; Length 2813;
Best Local Similarity 26.0%; Pred. No. 0.081;
Matches 19; Conservative 13; Mismatches 27; Indels 14; Gaps 2;

QY 8 GRG-----KCPSPNEIFSRCDGRQCFPCNV-VPKPLCIKICAPGVCRLGYL 53
Db 635 GRGVRVAMREPGRCELNCPKGVYLCQGTFCNLTCRSLSYPDCEACLEGCFPPGLY 694

QY 54 RNKKKVCVPRSKC 66
Db 695 MDERGDCVPKQAQC 707

RESULT 13
US-10-021-660-79
; Sequence 79, Application US/10021660
; Publication No. US20030152926A1
; GENERAL INFORMATION:
; APPLICANT: Murray, Richard
; APPLICANT: Glynn, Richard
```

```
; APPLICANT: Watson, Susan R.
; APPLICANT: EOS Biotechnology, Inc.
; TITLE OF INVENTION: No. US20030152926A1e1 Methods of Diagnosis of Angiogenesis,
; TITLE OF INVENTION: Compositions and Methods of Screening for Angiogenesis
; TITLE OF INVENTION: Modulators
; FILE REFERENCE: 018501-0007100S
; CURRENT APPLICATION NUMBER: US/10/021,660
; CURRENT FILING DATE: 2001-12-06
; PRIOR APPLICATION NUMBER: US/09/784,356
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: US 09/637,977
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 135
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 79
; LENGTH: 2813
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-021-660-79

Query Match      25.1%; Score 98; DB 12; Length 2813;
Best Local Similarity 26.0%; Pred. No. 0.081;
Matches 19; Conservative 13; Mismatches 27; Indels 14; Gaps 2;

QY 8 GRG-----KCPSPNEIFSRCDGRQCFPCNV-VPKPLCIKICAPGVCRLGYL 53
Db 635 GRGVRVAMREPGRCELNCPKGVYLCQGTFCNLTCRSLSYPDCEACLEGCFPPGLY 694

QY 54 RNKKKVCVPRSKC 66
Db 695 MDERGDCVPKQAQC 707

RESULT 14
US-10-020-141-4
; Sequence 4, Application US/10020141
; Publication No. US20030092013A1
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Jeanette
; APPLICANT: Ableson, Allen
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF VASCULAR DISEASE
; FILE REFERENCE: MMI-002
; CURRENT APPLICATION NUMBER: US/10/020,141
; CURRENT FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/313,097
; PRIOR FILING DATE: 2001-08-16
; PRIOR APPLICATION NUMBER: US 60/327,485
; PRIOR FILING DATE: 2001-10-05
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 2813
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-020-141-4

Query Match      25.1%; Score 98; DB 15; Length 2813;
Best Local Similarity 26.0%; Pred. No. 0.081;
Matches 19; Conservative 13; Mismatches 27; Indels 14; Gaps 2;

QY 8 GRG-----KCPSPNEIFSRCDGRQCFPCNV-VPKPLCIKICAPGVCRLGYL 53
Db 635 GRGVRVAMREPGRCELNCPKGVYLCQGTFCNLTCRSLSYPDCEACLEGCFPPGLY 694

QY 54 RNKKKVCVPRSKC 66
Db 695 MDERGDCVPKQAQC 707

RESULT 15
US-09-498-272-25
; Sequence 25, Application US/09498272
; Publication No. US20030113890A1
```

GENERAL INFORMATION:
APPLICANT: Vlasuk, George Phillip
Stanssens, Patrick Eric Hugo
Messens, Joris Hilda Lieven
Lauwereys, Marc Josef
Laroche, Yves Rene
Jespers, Laurent Stephane
Ganshans, Yannick Georges Jozef
Moyle, Matthew
Bergum, Peter W.
TITLE OF INVENTION: NEMATODE-EXTRACTED SERINE PROTEASE
INHIBITORS AND ANTICOAGULANT
PROTEIN
NUMBER OF SEQUENCES: 356
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/498,272
FILING DATE: 04-Feb-2000
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/13231
FILING DATE: October 17, 1995
APPLICATION NUMBER: 08/486,399
FILING DATE: June 5, 1995
APPLICATION NUMBER: 08/486,397
FILING DATE: June 5, 1995
APPLICATION NUMBER: 08/485,380
FILING DATE: June 5, 1995
APPLICATION NUMBER: 08/461,965
FILING DATE: June 5, 1995
APPLICATION NUMBER: 08/326,110
FILING DATE: October 18, 1994
ATTORNEY/AGENT INFORMATION:
NAME: BIGGS, SUZANNE L.
REGISTRATION NUMBER: 30,158
REFERENCE/DOCKET NUMBER: 216/270
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 82 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Ancylostoma ceylanicum
SEQUENCE DESCRIPTION: SEQ ID NO: 25:

US-09-498-272-25

Query Match 23.3%; Score 91; DB 11; Length 82;
Best Local Similarity 36.4%; Pred. No. 0.016;
Matches 24; Conservative 6; Mismatches 22; Indels 14; Gaps 4;

QY 12 CPSNEIFSRG--DGRQRFQ-----PNNVFKPLCKICAPG-CVCRIGYLRNKKKYC 60

DB 4 CGSNERYSGNDKQCKRKNEDDYKGDGACRSHVCER---PGACVCDGDFYRNKKKGC 60

QY 61 VPRSKC 66

Db 61 VESDDC 66

Search completed: November 17, 2003, 08:14:48
Job time : 30 secs

PT individual at risk for bee venom hypersensitivity -

PS Example 2; Page 26; 32pp; English.

XX The present sequence is that of one the four isoforms of Api m 6,
 CC designated Api m 6.01. The sequence represents the central amino acid
 CC sequence shared by all four isoforms (see ABB08331, ABB08332 and
 CC ABB08333). The specification describes a substantially pure polypeptide,
 CC Api m 6, derived from bee venom and found in four isoforms. The proteins
 CC of the invention have immunosuppressant activity and may form the basis
 CC as an allergen for immunotherapy. The protein is useful for identifying
 CC an individual at risk for bee venom hypersensitivity. The method
 CC comprises administering Api m 6 to the individual and measuring an immune
 CC response raised, where a detectable immune response indicates that the
 CC individual is at risk for bee venom hypersensitivity. Antibodies specific
 CC for Api m 6 are useful for purifying the protein.

SQ Sequence 67 AA;

Query Match 100.0%; Score 391; DB 23; Length 67;
 Best Local Similarity 100.0%; Pred. No. 8.9e-32;
 Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGFGLGGRGKCPSEIFSRCDGRCQRFPCPNVVPKLCIKICAPGCVCLGLYLNKKKVC 60
 Db 1 GGFGLGGRGKCPSEIFSRCDGRCQRFPCPNVVPKLCIKICAPGCVCLGLYLNKKKVC 60

Qy 61 VPRSKCG 67
 Db 61 VPRSKCG 67

RESULT 2

ABB08331

ID ABB08331 standard; protein; 69 AA.

AC ABB08331;

XX 18-JUN-2002 (first entry)

DE Bee venom protein Api m 6.02.

XX Bee venom; isoform; immunosuppressant; vaccine; Api m 6; immune response;
 KW bee venom hypersensitivity; antibody; protein purification; Api m 6.02;
 KW immunotherapy; allergen.

XX Apis sp.

OS WO200188085-A2.

XX 22-NOV-2001.

XX 16-FEB-2001; 2001WO-IB01736.

XX 18-FEB-2000; 2000US-0506978.

XX (ECOL-) ECOLE POLYTECHNIQUE FEDERALE LAUSANNE.

PI Spertini F;

XX WPI; 2002-082988/11.

XX New bee venom polypeptides, useful for modulating immune responses e.g.
 PT in individual hypersensitive to the venom and for identifying
 PT individual at risk for bee venom hypersensitivity -

PS Example 2; Page 26; 32pp; English.

XX The present sequence is that of one the four isoforms of Api m 6,
 CC designated Api m 6.02. All four isoforms share a common central amino
 CC acid sequence shared by all four isoforms (see ABB08330, ABB08332 and
 CC ABB08333). The specification describes a substantially pure polypeptide,
 CC Api m 6, derived from bee venom and found in four isoforms. The proteins
 CC of the invention have immunosuppressant activity and may form the basis
 CC as an allergen for immunotherapy. The protein is useful for identifying
 CC an individual at risk for bee venom hypersensitivity. The method
 CC comprises administering Api m 6 to the individual and measuring an immune
 CC response raised, where a detectable immune response indicates that the
 CC individual is at risk for bee venom hypersensitivity. Antibodies specific
 CC for Api m 6 are useful for purifying the protein.

SQ Sequence 69 AA;

Query Match 100.0%; Score 391; DB 23; Length 69;
 Best Local Similarity 100.0%; Pred. No. 9.1e-32;
 Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGFGLGGRGKCPSEIFSRCDGRCQRFPCPNVVPKLCIKICAPGCVCLGLYLNKKKVC 60
 Db 1 GGFGLGGRGKCPSEIFSRCDGRCQRFPCPNVVPKLCIKICAPGCVCLGLYLNKKKVC 60

Qy 61 VPRSKCG 67
 Db 61 VPRSKCG 67

RESULT 3

ABB08332

ID ABB08332 standard; protein; 71 AA.

AC ABB08332;

XX 18-JUN-2002 (first entry)

DE Bee venom protein Api m 6.03.

XX Bee venom; isoform; immunosuppressant; vaccine; Api m 6; immune response;
 KW bee venom hypersensitivity; antibody; protein purification; Api m 6.03;
 KW immunotherapy; allergen.

XX Apis sp.

OS WO200188085-A2.

XX 22-NOV-2001.

XX 16-FEB-2001; 2001WO-IB01736.

XX 18-FEB-2000; 2000US-0506978.

XX (ECOL-) ECOLE POLYTECHNIQUE FEDERALE LAUSANNE.

PI Spertini F;

XX WPI; 2002-082988/11.

XX New bee venom polypeptides, useful for modulating immune responses e.g.
 PT in individual hypersensitive to the venom and for identifying
 PT individual at risk for bee venom hypersensitivity -

PS Example 2; Page 26; 32pp; English.

XX The present sequence is that of one the four isoforms of Api m 6,
 CC designated Api m 6.03. All four isoforms share a common central amino
 CC acid sequence shared by all four isoforms (see ABB08330, ABB08331 and
 CC ABB08333). The specification describes a substantially pure polypeptide,
 CC Api m 6, derived from bee venom and found in four isoforms. The proteins
 CC of the invention have immunosuppressant activity and may form the basis
 CC of the invention have immunosuppressant activity and may form the basis

CC of a vaccine. Api m 6 is useful for modulating an immune response, i.e.
CC as an allergen for immunotherapy. The protein is useful for identifying
CC an individual at risk for bee venom hypersensitivity. The method
CC comprises administering Api m 6 to the individual and measuring an immune
CC response raised, where a detectable immune response indicates that the
CC individual is at risk for bee venom hypersensitivity. Antibodies specific
CC for Abi m 6 are useful for purifying the protein.

XX
SQ Sequence 71 AA;
Query Match 100.0%; Score 391; DB 23; Length 71;
Best Local Similarity 100.0%; Pred. No. 9.3e-32;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGFGLGGRGKCPSEIFSRCDGRQFCFNNVVKPLCTIKICAPGCVCLGVLNRNKKVC 60
|||||
Db 5 GGFGLGGRGKCPSEIFSRCDGRQFCFNNVVKPLCTIKICAPGCVCLGVLNRNKKVC 64
QY 61 VPRSKCG 67
|||||
Db 65 VPRSKCG 71

RESULT 4
ABB08333
ID ABB08333 standard; protein; 73 AA.
AC ABB08333;
XX
DT 18-JUN-2002 (first entry)
XX
DE Bee venom protein Api m 6.04.
XX
KW Bee venom; isoform; immunosuppressant; vaccine; Api m 6; immune response;
KW bee venom hypersensitivity; antibody; protein purification; Api m 6.04;
KW immunotherapy; allergen.

XX
OS Apis sp.
XX
FH Key Location/Qualifiers
FT Misc-difference 72 /label= Pro, Leu
FT FT Misc-difference 73 /label= Pro, Leu
FT FT
XX WO200188085-A2.
PN
XX
PD 22-NOV-2001.
XX
PF 16-FEB-2001; 2001WO-IB01736.
XX
PR 18-FEB-2000; 2000US-0506978.
XX
PA (ECOL-) ECOLE POLYTECHNIQUE FEDERALE LAUSANNE.
XX
PI Spertini F;
XX
DR WPI; 2002-082988/11.
XX
PT New bee venom polypeptides, useful for modulating immune responses e.g.
PT in individual hypersensitive to the venom and for identifying
PT individual at risk for bee venom hypersensitivity -
XX
FS Example 2; Page 26; 32pp; English.

XX The present sequence is that of one the four isoforms of Api m 6.
CC designated Api m 6.04. All four isoforms share a common central amino
CC acid sequence shared by all four isoforms (see ABB08330, ABB08331 and
CC ABB08332). The specification describes a substantially pure polypeptide,
CC Api m 6, derived from bee venom and found in four isoforms. The proteins
CC of the invention have immunosuppressant activity and may form the basis
CC of a vaccine. Api m 6 is useful for modulating an immune response, i.e.
CC as an allergen for immunotherapy. The protein is useful for identifying

CC an individual at risk for bee venom hypersensitivity. The method
CC comprises administering Api m 6 to the individual and measuring an immune
CC response raised, where a detectable immune response indicates that the
CC individual is at risk for bee venom hypersensitivity. Antibodies specific
CC for Abi m 6 are useful for purifying the protein.

XX
SQ Sequence 73 AA;
Query Match 100.0%; Score 391; DB 23; Length 73;
Best Local Similarity 100.0%; Pred. No. 9.6e-32;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGFGLGGRGKCPSEIFSRCDGRQFCFNNVVKPLCTIKICAPGCVCLGVLNRNKKVC 60
|||||
Db 5 GGFGLGGRGKCPSEIFSRCDGRQFCFNNVVKPLCTIKICAPGCVCLGVLNRNKKVC 64
QY 61 VPRSKCG 67
|||||
Db 65 VPRSKCG 71

RESULT 5
AAI69209
ID AAI69209 standard; Protein; 92 AA.
XX
AC AAI69209;
XX
DT 30-MAY-2000 (first entry)
XX
DE Amino acid sequence of honey bee venom PX3.101 protein.
XX
KW Protein PX3.101; honey bee; venom; interleukin-8; IL-8; receptor;
KW CXCR1; CXCR2; cyclooxygenase; lipoxigenase; phospholipase; protease;
KW inflammatory disease; gene therapy; cancer; autoimmune disease; pain;
KW psoriasis; systemic lupus erythematosus; multiple sclerosis;
KW scleroderma; metastatic cancer; Alzheimer's disease; vasculitis;
KW aging process; antigen.

XX
OS Apis mellifera.
XX
FH Key Location/Qualifiers
FT Peptide 1..19 /note= "signal peptide"
FT FT
FT Region 20..34 /note= "this region contains 5 GGG repeats"
XX
PN GB2341389-A.
XX
PD 15-MAR-2000.
XX
PF 13-SEP-1999; 99GB-0021605.
XX
PR 14-SEP-1998; 98US-0100172.
XX
PA (PANP-) PAN PACIFIC PHARM INC.
XX
PI Chi X, Lu Y;
XX
DR WPI; 2000-185368/17.
DR N-PSDB; AA261247.
XX
PT Isolated nucleic acids encoding the bee venom protein PX3.101 useful
PT for treating autoimmune and inflammatory disorders such as rheumatoid
PT arthritis -
XX
PS Claim 2; Fig 3A-B; 83pp; English.

XX The present sequence represents the protein PX3.101, which is a honey
CC bee venom isolated Apis mellifera. PX3.101 inhibits the binding of
CC interleukin-8 (IL-8) to its receptor (e.g. CXCR1 and CXCR2) and
CC inhibits a variety of enzymes (e.g. cyclooxygenases, lipoxigenases,
CC phospholipases and proteases) associated with inflammatory diseases.

CC The nucleic acids may be used for the recombinant production of
CC PX3.101 proteins either in vivo (as part of a gene therapy protocol)
CC or in vitro (as a fermentation culture). The nucleic acids may also
CC be used as probes to identify similar sequences in samples. The PX3.101
CC protein may be used for the treatment of inflammatory diseases, cancers,
CC autoimmune diseases, pain and/or diseases associated with chemokine
CC (especially IL-8) imbalances such as rheumatoid arthritis, multiple
CC sclerosis, psoriasis, systemic lupus erythematosus (SLE), Crohn's
CC disease, vasculitis, scleroderma, metastatic cancer and Alzheimer's
CC disease in humans. It is also disclosed that the proteins may be used
CC to accelerate wound healing, reduce several aging processes and protect
CC against ultraviolet light. The proteins may also be used as antigens in
CC the production of antibodies specific for PX3.101. The antibodies may
CC be used as diagnostic agents to detect PX3.101 protein in samples and
CC to down regulate PX3.101 activity.

XX
SQ Sequence 92 AA;

Query Match 100.0%; Score 391; DB 21; Length 92;
Best Local Similarity 100.0%; Pred. No. 1.2e-31;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFGGLGGRGKCPGNEIFSRDGRQFCFNVVVKPLCIKICAPGCVCRLLGRLNKKVC 60
|||||
Db 26 GFGGLGGRGKCPGNEIFSRDGRQFCFNVVVKPLCIKICAPGCVCRLLGRLNKKVC 85
|||||

Qy 61 VPRSKCG 67
|||||

Db 86 VPRSKCG 92
|||||

RESULT 6
AB971529
ID ABB71529 standard; Protein; 3843 AA.
XX
AC ABB71529;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 41379.
DE
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX N-PSDB; ABL15632.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Disclosure; SEQ ID NO 41379; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 3843 AA;

Query Match 27.8%; Score 108; DB 22; Length 3843;
Best Local Similarity 33.9%; Pred. No. 0.041;
Matches 19; Conservative 7; Mismatches 30; Indels 0; Gaps 0;

Qy 11 KCPSENFISFSCDGRQFCFNVVVKPLCIKICAPGCVCRLLGRLNKKVCVPRSKC 66
|||||
Db 1130 KCPLGVDFDECGDGCALSCDDLFSKSGSKRECVGRCRPHGEVYNEDEGCVPRKMC 1185
|||||

RESULT 7
AAW67738
ID AAW67738 standard; Protein; 2476 AA.
XX
XX AAW67738;
XX
XX 16-MAR-1999 (first entry)
XX
XX Pig p105 zona pellucida-binding protein.
DE
XX Pig; porcine; sperm; egg-binding protein; zona pellucida; contraception;
XX fertilisation.
XX
XX Sus scrofa.
XX
XX US5851817-A.
XX
XX 22-DEC-1998.
XX
XX 19-JUL-1994; 94US-0276967.
XX
XX 19-JUL-1994; 94US-0276967.
XX
XX (TEXA) UNIV TEXAS SYSTEM.
XX
XX Garbers DL, Hardy DM;
XX
XX WPI; 1999-080410/07.
XX
XX N-PSDB; AAV81446.
XX
XX DNA encoding porcine sperm egg-binding protein - useful for
XX producing recombinant protein
XX
XX Claim 1; Fig 8A-B; 47pp; English.

XX This sequence represents a pig sperm egg-binding protein designated
XX protein p105. Porcine sperm proteins which bind the zona pellucida in a
XX species-specific manner were isolated. Separation by gel electrophoresis
XX resulted in bands of proteins with molecular weights of 130, 150 and
XX 170 kD, under native conditions but bands of 105 and 45 kD were observed
XX when the proteins were separated under denaturing conditions. Compounds
XX which bind these proteins can be used for contraception or fertilisation.

XX
SQ Sequence 2476 AA;

Query Match 27.5%; Score 107.5; DB 20; Length 2476;
Best Local Similarity 33.3%; Pred. No. 0.031;
Matches 22; Conservative 9; Mismatches 18; Indels 17; Gaps 3;

Qy 11 KCPSENFISFSCDGRQFCFNVVVKPLCIKICAPGCVCRLLGRLNKKVCV 61
|||||
Db 1455 KCPSSGSSYSTCANPCPATCLSLNNPSCYCPSTLP-----CABGCEQKHILSGTS-CV 1506
|||||

Qy 62 PRSKCG 67
|||||

XX WO9527057-A1.
XX PD 12-OCT-1995.
XX PF 03-APR-1995; 95WO-JP00638.
XX PR 30-MAR-1995; 95JP-0109927.
XX PR 01-APR-1994; 94JP-0129487.
XX PR 24-AUG-1994; 94JP-0222547.
XX PA (CHUS) CHUGAI SEIYAKU KK.
XX PI Harada N, Morikawa M;
XX DR WPI; 1995-358632/46.
XX DR N-PSDB; AAT63074.
XX PT DNA derived from colonic epithelium encoding IgG-Fc binding protein
XX PT - used in the mapping and analysis of IgG-Fc binding protein mRNA
XX PS Claim 3; Page 86-113; 132pp; Japanese.
XX CC This sequence represents the IgG-Fc binding protein of human colonic
XX CC epithelium. This sequence was isolated using the sequence given in
XX CC AAT63074. mRNA isolated from human colonic epithelial tissue was used
XX CC to prepare a cDNA library. This was screened using monoclonal
XX CC antibodies K9 and K17 which bind to the large and small components of the
XX CC binding protein. Active clones, see also AAT63077-81, were used to
XX CC derive probes for screening a second DNA library from human colonic
XX CC epithelial tissue.
XX SQ Sequence 5405 AA;
Query Match 27.2%; Score 106.5; DB 16; Length 5405;
Best Local Similarity 36.8%; Pred. No. 0.078;
Matches 21; Conservative 6; Mismatches 29; Indels 1; Gaps 1;
QY 11 KCPNNEIFSRCDGRCQRCFNPVVPKLCIKICAPGCVCLGVLNKKKVCVPRSKCG 67
Db 1531 ECPNNSHYELCADTCSLGSALSAPPQCDGCAEGCQCDGFLYN-GQACVPIQCG 1586
RESULT 11
ID ABP55383
AC ABP55383; Protein; 5405 AA.
XX DT 30-JAN-2003 (first entry)
XX DE Human colon tumour related protein sequence SEQ ID NO:1116.
XX KW Human; colon cancer; colon tumour; immunotherapy; diagnosis; cancer;
XX KW tumour; immune response; immunostimulant; cytostatic; vaccine.
XX OS Homo sapiens.
XX PN WO200283070-A2.
XX PD 24-OCT-2002.
XX PF 09-APR-2002; 2002WO-US11475.
XX PR 10-APR-2001; 2001US-0833263.
XX PR 03-AUG-2001; 2001US-0922217.
XX PR 19-DEC-2001; 2001US-0025380.
XX PA (CORI-) CORIXA CORP.
XX PI Xu J, Lodes MJ, Secretist H, Benson DR, Meagher MJ, Stolk JA;
XX PI Wang T, Jiang Y, Smith CL, King GE, Wang A, Clapper JD;
XX PI Skelky YAW, Fanger GR, Vedwick TS, Carter D;

XX WPI; 2003-067548/06.
XX DR N-PSDB; ABZ33718.
XX PT New polynucleotide, useful for the preparation of a composition for
XX PT stimulating an immune response against, or treating, cancer -
XX XX Example 8; Page 511-522; 537pp; English.
XX CC The present invention describes compounds (I) for the immunotherapy and
XX CC diagnosis of colon cancer. Also described: (1) a method for detecting
XX CC the presence of cancer in a patient; (2) a method for stimulating and/or
XX CC expanding T cells specific for a tumour protein; (3) an isolated T cell
XX CC population comprising T cells prepared by the method of (2); (4) a method
XX CC for stimulating an immune response in a patient; (5) a method for
XX CC treating cancer in a patient; and (6) a method for inhibiting the
XX CC development of cancer in a patient. (I) have immunostimulant and
XX CC cytostatic activities and can be used in vaccines. ABZ32646 to ABZ33725
XX CC and ABP55343 to ABP55391 represent human colon cancer/tumour related
XX CC sequences used in the exemplification of the present invention.
XX SQ Sequence 5405 AA;
Query Match 27.2%; Score 106.5; DB 24; Length 5405;
Best Local Similarity 36.8%; Pred. No. 0.078;
Matches 21; Conservative 6; Mismatches 29; Indels 1; Gaps 1;
QY 11 KCPNNEIFSRCDGRCQRCFNPVVPKLCIKICAPGCVCLGVLNKKKVCVPRSKCG 67
Db 1531 ECPNNSHYELCADTCSLGSALSAPPQCDGCAEGCQCDGFLYN-GQACVPIQCG 1586
RESULT 12
ID ABG22216
XX AC ABG22216; Protein; 7337 AA.
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #22207.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
XX DR N-PSDB; AAS86403.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity -
XX PS Claim 20; SEQ ID No 52575; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2003, 08:03:11 ; Search time 11 Seconds
(without alignments)
286.435 Million cell updates/sec

Title: US-09-506-978-1

Perfect score: 391

Sequence: 1 GGGGLGGGKCPSEIFSR.....CRLGYLRKKKVCVPSKCG 67

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	113	28.9	5376	1 ZAN_MOUSE	O88799 mus musculus
2	109	27.9	2812	1 ZAN_HUMAN	Q9Y493 homo sapien
3	107.5	27.5	2476	1 ZAN_PIG	O28983 sus scrofa
4	104.5	26.7	56	1 AMCT_APIME	P56682 apis mellif
5	103.5	26.5	2813	1 VWF_CANFA	Q28295 canis famil
6	101.5	26.0	937	1 VWF_BOVIN	P80012 bos taurus
7	98.5	25.2	2282	1 ZAN_RABIT	P57999 oryctolagus
8	98	25.1	115	1 A62F_DROME	O46202 drosophila
9	98	25.1	2813	1 VWF_HUMAN	P04275 homo sapien
10	97	24.8	77	1 CVP5_PIMHY	Q8t0w0 pimply hypo
11	91.5	23.4	2482	1 VWF_PIG	Q28833 sus scrofa
12	90.5	23.1	3133	1 HMCT_BOMMO	P38032 bombyx mori
13	86	22.0	1700	1 BAR3_CHITE	Q03376 chironomus
14	84.5	21.6	1513	1 MUC2_RAT	Q62635 rattus norv
15	80.5	20.6	63	1 ICE1_ASCSU	P07851 ascaris suu
16	80.5	20.6	99	1 VOJ2_CABEL	P34625 caenorhabdi
17	80	20.5	1370	1 IG1R_RAT	P24052 rattus norv
18	80	20.5	1373	1 IG1R_MOUSE	Q60751 mus musculu
19	79.5	20.3	351	1 CRMB_COMPX	P07559 cowpox viru
20	78.5	20.1	65	1 ICE2_ASCSU	P07852 ascaris suu
21	76.5	19.6	72	1 MT1A_MYTED	P80246 mytilus edu
22	76.5	19.6	72	1 MT1B_MYTED	O62554 mytilus edu
23	76.5	19.6	349	1 CRMB_CAMPS	Q8uya7 camelpox vi
24	76.5	19.6	855	1 STL4_MOUSE	P36677 mus musculu
25	75.5	19.3	349	1 CRMB_VARV	P34015 variola vir
26	75.5	19.3	5179	1 MUC2_HUMAN	Q02817 homo sapien
27	75	19.2	956	1 TSP3_MOUSE	Q05895 mus musculu
28	75	19.2	4289	1 TENX_HUMAN	P22105 homo sapien
29	73	18.7	60	1 MT_PERFL	P22725 perca fluvi
30	72.5	18.5	1416	1 YN81_CABEL	Q03610 caenorhabdi
31	72	18.4	326	1 VT2_MYXVL	P29825 myxoma viru
32	72	18.4	956	1 TSP3_HUMAN	P49746 homo sapien
33	72	18.4	1367	1 IG1R_HUMAN	P08069 homo sapien

ALIGNMENTS

RESULT 1

ZAN_MOUSE

ID	ZAN_MOUSE	STANDARD;	PRT;	5376 AA.
AC	O88799; O08647;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Zonadhesin precursor.			
GN	ZAN.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Testis;			
RX	MEDLINE=98123114; PubMed=9452463;			
RA	Gao Z., Garbers D.L.;			
RT	"Species diversity in the structure of zonadhesin, a sperm-specific membrane protein containing multiple cell adhesion molecule-like domains".			
RT	J. Biol. Chem. 273:3415-3421(1998).			
RN	[2]			
RP	SEQUENCE OF 4864-5376 FROM N.A.			
RC	TISSUE=Testis;			
RX	MEDLINE=97271566; PubMed=9126492;			
RA	Gao Z., Harumi T., Garbers D.L.;			
RT	"Chromosome localization of the mouse zonadhesin gene and the human zonadhesin gene (ZAN)".			
RL	Genomics 41:119-122(1997).			
CC	-1- FUNCTION: BINDS IN A SPECIES-SPECIFIC MANNER TO THE ZONA PELLUCIDA OF THE EGG. MAY BE INVOLVED IN GAMETE RECOGNITION AND/OR SIGNALING.			
CC	-1- SUBUNIT: PROBABLY FORMS COVALENT OLIGOMERS.			
CC	-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, EXCLUSIVELY ON THE APICAL REGION OF THE SPERM HEAD.			
CC	-1- TISSUE SPECIFICITY: IN TESTIS; PRIMARILY IN HAPLOID SPERMATIDS.			
CC	-1- DOMAIN: THE MAM DOMAINS PROBABLY MEDIATES SPERM ADHESION TO THE ZONA PELLUCIDA.			
CC	-1- DOMAIN: DURING SPERM MIGRATION THROUGH THE REPRODUCTIVE TRACTS, THE MUCIN-LIKE DOMAIN MIGHT INHIBIT INAPPROPRIATE TRAPPING OF SPERMATOZOEA OR PROMOTING ADHESION TO THE OVULATORY ISTHMUS.			
CC	-1- DOMAIN: THE WFP DOMAIN 2 MAY MEDIATE COVALENT OLIGOMERIZATION (BY SIMILARITY TO HUMAN INTESTINAL MUCIN MUC2).			
CC	-1- SIMILARITY: Contains 3 MAM domains.			
CC	-1- SIMILARITY: Contains 25 WFP domains.			
CC	-1- SIMILARITY: Contains 1 EGF-like domain.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).			
CC	-----			

```

DR EMBL; U97068; AAC26680.1; -.
DR EMBL; U83190; AAC53125.1; -.
DR PIR; T42215; T42215.
DR MGI; MGI:106656; ZAN.
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR003645; FcI_N.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR000998; MAM domain.
DR InterPro; IPR002919; TIL_Cysrich.
DR InterPro; IPR003328; TILA_Cysrich.
DR InterPro; IPR001007; VWF_C.
DR InterPro; IPR001846; VWF_D.
DR Pfam; PF00629; MAM; 3.
DR Pfam; PF01826; TIL; 25.
DR Pfam; PF02345; TILA; 25.
DR Pfam; PF00094; vwd; 4.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00274; FcI_N; 11.
DR SMART; SM00137; MAM; 2.
DR SMART; SM00214; VWC; 17.
DR SMART; SM00216; VWD; 4.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 18.
DR PROSITE; PS00740; MAM_1; FALSE_NEG.
DR PROSITE; PS00660; MAM_2; 3.
KW Signal; Glycoprotein; Transmembrane; EGF-like domain; Cell adhesion;
KW Repeat.
FT SIGNAL 1 17
FT CHAIN 18 5376
FT DOMAIN 18 5310
FT TRANSMEM 5311 5337
FT DOMAIN 5338 5376
FT DOMAIN 45 210
FT DOMAIN 215 374
FT DOMAIN 377 542
FT DOMAIN 547 1170
FT DOMAIN 1171 1280
FT DOMAIN 1281 1669
FT DOMAIN 1670 2056
FT DOMAIN 2057 2459
FT DOMAIN 2460 2579
FT DOMAIN 2580 2699
FT DOMAIN 2700 2819
FT DOMAIN 2820 2939
FT DOMAIN 2940 3059
FT DOMAIN 3060 3179
FT DOMAIN 3180 3299
FT DOMAIN 3300 3416
FT DOMAIN 3417 3536
FT DOMAIN 3537 3656
FT DOMAIN 3657 3776
FT DOMAIN 3777 3892
FT DOMAIN 3893 4029
FT DOMAIN 4029 4148
FT DOMAIN 4149 4263
FT DOMAIN 4264 4383
FT DOMAIN 4384 4503
FT DOMAIN 4504 4623
FT DOMAIN 4624 4743
FT DOMAIN 4744 4863
FT DOMAIN 4864 5261
FT DOMAIN 5259 5295
FT DISULFID 5263 5274
FT DISULFID 5268 5283
FT DISULFID 5285 5294
FT CARBOHYD 339 339
FT CARBOHYD 499 499
FT CARBOHYD 1216 1216
FT CARBOHYD 1239 1239
FT CARBOHYD 1314 1314
FT CARBOHYD 1814 1814
FT CARBOHYD 1908 1908
DR EMBL; U97068; AAC26680.1; -.
DR EMBL; U83190; AAC53125.1; -.
DR PIR; T42215; T42215.
DR MGI; MGI:106656; ZAN.
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR003645; FcI_N.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR000998; MAM domain.
DR InterPro; IPR002919; TIL_Cysrich.
DR InterPro; IPR003328; TILA_Cysrich.
DR InterPro; IPR001007; VWF_C.
DR InterPro; IPR001846; VWF_D.
DR Pfam; PF00629; MAM; 3.
DR Pfam; PF01826; TIL; 25.
DR Pfam; PF02345; TILA; 25.
DR Pfam; PF00094; vwd; 4.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00274; FcI_N; 11.
DR SMART; SM00137; MAM; 2.
DR SMART; SM00214; VWC; 17.
DR SMART; SM00216; VWD; 4.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 18.
DR PROSITE; PS00740; MAM_1; FALSE_NEG.
DR PROSITE; PS00660; MAM_2; 3.
KW Signal; Glycoprotein; Transmembrane; EGF-like domain; Cell adhesion;
KW Repeat.
FT SIGNAL 1 17
FT CHAIN 18 5376
FT DOMAIN 18 5310
FT TRANSMEM 5311 5337
FT DOMAIN 5338 5376
FT DOMAIN 45 210
FT DOMAIN 215 374
FT DOMAIN 377 542
FT DOMAIN 547 1170
FT DOMAIN 1171 1280
FT DOMAIN 1281 1669
FT DOMAIN 1670 2056
FT DOMAIN 2057 2459
FT DOMAIN 2460 2579
FT DOMAIN 2580 2699
FT DOMAIN 2700 2819
FT DOMAIN 2820 2939
FT DOMAIN 2940 3059
FT DOMAIN 3060 3179
FT DOMAIN 3180 3299
FT DOMAIN 3300 3416
FT DOMAIN 3417 3536
FT DOMAIN 3537 3656
FT DOMAIN 3657 3776
FT DOMAIN 3777 3892
FT DOMAIN 3893 4029
FT DOMAIN 4029 4148
FT DOMAIN 4149 4263
FT DOMAIN 4264 4383
FT DOMAIN 4384 4503
FT DOMAIN 4504 4623
FT DOMAIN 4624 4743
FT DOMAIN 4744 4863
FT DOMAIN 4864 5261
FT DOMAIN 5259 5295
FT DISULFID 5263 5274
FT DISULFID 5268 5283
FT DISULFID 5285 5294
FT CARBOHYD 339 339
FT CARBOHYD 499 499
FT CARBOHYD 1216 1216
FT CARBOHYD 1239 1239
FT CARBOHYD 1314 1314
FT CARBOHYD 1814 1814
FT CARBOHYD 1908 1908

```

RA Chéung T.L., Hardy D.M., Schwartz S., Scherer S.W., Tsui L.-C.,
RA Miller W., Koop B.F.;
RT "Comparative analysis of the gene-dense ACHE/TER2 region on human
RT chromosome 7q22 with the orthologous region on mouse chromosome 5.";
RL Nucleic Acids Res. 29:1352-1365(2001).
RN [4]
RP SEQUENCE OF 2375-2683 FROM N.A. (ISOFORM 7).
RC TISSUE=Testis;
RX MEDLINE=97271566; PubMed=9126492;
RA Gao Z., Harumi T., Garbers D.L.;
RT zonadhesin gene (zan).";
RL "Chromosome localization of the mouse zonadhesin gene and the human
RL zonadhesin gene (zan).";
RL Genomics 41:119-122(1997).
CC -!- FUNCTION: BINDS IN A SPECIES-SPECIFIC MANNER TO THE ZONA PELLUCIDA
CC OF THE EGG. MAY BE INVOLVED IN GAMETE RECOGNITION AND/OR
CC SIGNALING.
CC -!- SUBUNIT: PROBABLY FORMS COVALENT OLIGOMERS.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, EXCLUSIVELY ON THE
CC APICAL REGION OF THE SPERM HEAD (BY SIMILARITY).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=7;
CC Name=3;
CC IsoId=Q9Y493-1; Sequence=Displayed;
CC Name=1;
CC IsoId=Q9Y493-2; Sequence=VSP_001430, VSP_001431;
CC Name=2;
CC IsoId=Q9Y493-3; Sequence=VSP_001428, VSP_001429;
CC Name=4;
CC IsoId=Q9Y493-4; Sequence=VSP_001424, VSP_001425;
CC Name=5;
CC IsoId=Q9Y493-5; Sequence=VSP_001420, VSP_001421;
CC Name=6;
CC IsoId=Q9Y493-6; Sequence=VSP_001422, VSP_001423;
CC Name=7;
CC IsoId=Q9Y493-7; Sequence=VSP_001426, VSP_001427;
CC -!- TISSUE SPECIFICITY: IN TESTIS; PRIMARILY IN HAPLOID SPERMATIDS.
CC -!- DOMAIN: THE MAM DOMAINS PROBABLY MEDIATE SPERM ADHESION TO THE
CC ZONA PELLUCIDA.
CC -!- DOMAIN: DURING SPERM MIGRATION THROUGH THE REPRODUCTIVE TRACTS,
CC THE MUCIN-LIKE DOMAIN MIGHT INHIBIT INAPPROPRIATE TRAPPING OF
CC SPERMATOZOA OR PROMOTING ADHESION TO THE OVIDUCTAL ISTHMUS.
CC -!- DOMAIN: The VWF domain 2 may mediate covalent oligomerization (by
CC similarity to human intestinal mucin MUC2).
CC -!- SIMILARITY: Contains 3 MAM domains.
CC -!- SIMILARITY: Contains 4 VWF domains.
CC -!- SIMILARITY: Contains 1 EGF-like domain.
CC -!- CAUTION: Ref.2 sequence differs from that shown due to
CC transposition of a number of exons.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF332975; AAK01431.1; -
DR EMBL; AF332976; AAK01432.1; -
DR EMBL; AF332977; AAK01433.1; -
DR EMBL; AF332978; AAK01434.1; -
DR EMBL; AF332979; AAK01435.1; -
DR EMBL; AF332980; AAK01436.1; -
DR EMBL; AY046055; AAL04410.1; -
DR EMBL; AY046055; AAL04411.1; -
DR EMBL; AY046055; AAL04412.1; -
DR EMBL; AY046055; AAL04413.1; -
DR EMBL; AY046055; AAL04414.1; -
DR EMBL; AY046055; AAL04415.1; -
DR EMBL; AF053356; AAC78790.1; ALT_SEQ.
DR EMBL; AF312032; AAC21011.1; -
DR EMBL; U83191; AAC51208.1; -
DR Genew; HGNC:12857; ZAN.

DR MIM; 602372; -
DR GO; GO:0016324; C:apical plasma membrane; NAS.
DR GO; GO:0007339; P:binding of sperm to zona pellucida; NAS.
DR GO; GO:0008037; P:cell recognition; NAS.
DR GO; GO:0016337; P:cell-cell adhesion; NAS.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR000998; MAM_Domain.
DR InterPro; IPR003328; TIL_Cysrich.
DR InterPro; IPR001007; VWF_C.
DR InterPro; IPR001846; VWF_D.
DR Pfam; PF00629; MAM; 3.
DR Pfam; PF01826; TIL; 4.
DR Pfam; PF02345; TILa; 5.
DR Pfam; PF00094; vwd; 4.
DR SMART; SM00181; EGF; 4.
DR SMART; SM00137; MAM; 3.
DR SMART; SM00214; VWC; 4.
DR SMART; SM00215; VWC_out; 4.
DR SMART; SM00216; VWD; 4.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 4.
DR PROSITE; PS00740; MAM_1; 1.
DR PROSITE; PS00600; MAM_2; 3.
DR Signal; Glycoprotein; Transmembrane; EGF-like domain; Cell adhesion;
KW Repeat; Alternative splicing.
FT SIGNAL 1 17
FT CHAIN 18 2812
FT ZONADHESIN.
FT DOMAIN 18 2757
FT EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 2758 2778
FT POTENTIAL.
FT DOMAIN 2779 2812
FT CYTOPLASMIC (POTENTIAL).
FT DOMAIN 39 204
FT MAM 1.
FT DOMAIN 209 368
FT MAM 2.
FT DOMAIN 371 536
FT MAM 3.
FT DOMAIN 573 1041
FT 66 X HEPTAPEPTIDE REPEATS (APPROXIMATE)
FT (MUCIN-LIKE DOMAIN).
FT DOMAIN 1156 1308
FT VWF 1.
FT DOMAIN 1542 1697
FT VWF 2.
FT DOMAIN 1931 2086
FT VWF 3.
FT DOMAIN 2331 2485
FT VWF 4.
FT DOMAIN 2708 2744
FT EGF-LIKE.
FT DISULFID 2712 2732
FT BY SIMILARITY.
FT DISULFID 2717 2732
FT BY SIMILARITY.
FT DISULFID 2734 2743
FT BY SIMILARITY.
FT CARBOHYD 333 333
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 493 493
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1112 1112
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1188 1188
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1685 1685
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1804 1804
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1900 1900
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1946 1946
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2203 2203
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2542 2542
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2701 2701
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 2597 2601
FT HGVS -> WAAAP (in isoform 5).
FT FTId=VSP_001420.
FT Missing (in isoform 5).
FT VARSPLIC 2602 2812
FT FTId=VSP_001421.
FT HGVSRYHISELYDTLPILC -> YAILCQEAAGALAGWR
FT VARSPLIC 2597 2617
FT DRTL (in isoform 6).
FT FTId=VSP_001422.
FT Missing (in isoform 6).
FT VARSPLIC 2618 2708
FT FTId=VSP_001423.
FT HGVSRYHISELYDTLPILCQQRPRG -> YAILCQEAAG
FT VARSPLIC 2597 2624
FT AALAGWRDRLCAGQQLSD (in isoform 4).
FT FTId=VSP_001424.
FT Missing (in isoform 4).
FT VARSPLIC 2625 2812
FT FTId=VSP_001425.
FT HGVSRYHISELYDTLPILCQQRPRGRLGRRLRQH
FT VARSPLIC 2597 2636
FT -> YAILCQEAAGALAGWRDRLCAMECPAGTIYQCMTPC
FT PASCANLADPGCEGPCVREGCAD (in isoform 7).
FT FTId=VSP_001426.


```
FT CARBOHYD 2122 2122 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2165 2165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2178 2178 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2329 2329 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2359 2359 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 823 823 C -> V (IN REF. 1; AA SEQUENCE).
FT CONFLICT 923 923 S -> Y (IN REF. 1; AA SEQUENCE).
FT CONFLICT 965 965 W -> Y (IN REF. 1; AA SEQUENCE).
FT CONFLICT 1241 1241 S -> K (IN REF. 1; AA SEQUENCE).
SQ SEQUENCE 2476 AA; 270364 MW; A13B690375A6548C CRC64;

Query Match
Best Local Similarity 33.3%; Pred. No. 0.00068;
Matches 22; Conservative 9; Mismatches 18; Indels 17; Gaps 3;

QY 11 KCPSEIFSRCDGRC-----QRFPCNVVPKPLCIKICAPGCVRLGYLRNKKVCV 61
Db 1455 KCPSSGYSYTCANPCPCATCLSLNNPSYCPSTLP-----CAREGCEQKHILSGTS-CV 1506

QY 62 PRSKCG 67
Db 1507 FLSQCG 1512

RESULT 4
AMCI APIME
ID _AMCI APIME STANDARD; PRT; 56 AA.
AC P56682;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE Chymotrypsin inhibitor (AMCI).
OS Apis mellifera (Honeybee).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea;
OC Apidae; Apis.
OX NCBI_TaxID=7460;
RN [1]
RP SEQUENCE AND STRUCTURE BY NMR.
RC TISSUE=Hemolymph;
RX MEDLINE=99339935; PubMed=10411628;
RA Bania J., Stachowiak D., Polanowski A.;
RT "Primary structure and properties of the cathepsin G/chymotrypsin
inhibitor from the larval hemolymph of Apis mellifera.";
RL Eur. J. Biochem. 262:680-687(1999).
CC -!- FUNCTION: Chymotrypsin and cathepsin G inhibitor.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Contains 1 TIL (Trypsin inhibitory-like) domain.
DR PDB; 1CCV; 12-MAR-99.
DR InterPro; IPR002919; TIL_Cysrich.
DR Pfam; PF01826; TIL; 1.
KW Serine protease inhibitor; 3D-structure.
FT DOMAIN 3 56 TIL.
FT DISULFID 3 36
FT DISULFID 12 32
FT DISULFID 16 28
FT DISULFID 20 56
FT DISULFID 38 50
FT TURN 5 6
FT TURN 7 13
FT STRAND 15 16
FT TURN 33 38
FT STRAND 40 41
FT TURN 43 45
FT TURN 46 48
FT STRAND 49 51
FT TURN 53 56
FT HELIX 53 56
SQ SEQUENCE 56 AA; 5973 MW; 092B2815AE6B2B7F CRC64;

Query Match
Best Local Similarity 36.8%; Pred. No. 5.8e-05;
Matches 21; Conservative 7; Mismatches 26; Indels 3; Gaps 2;
```

```
QY 11 KCPSEIFSRCDGRCQFPCNVVPKP-LCIKICAPGCVRLGYLRNKKVCVPRSKC 66
Db 2 ECGPNEVFNTGSGACAPTCAQ--PKTRICTMQCRIGCCQCGFLRNGEGACVLPENC 56

RESULT 5
VWF CANFA
ID _VWF CANFA STANDARD; PRT; 2813 AA.
AC Q28295; Q28311; Q9TS14;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Von Willebrand factor precursor (VWF).
GN VWF OR F8VWF.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Stoy S.J., Shibuya H., Nonneman D.J., Holzhauser J., Mohammed I.H.,
RA Johnson G.S.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Montgomery R.R., Fahs S., Montgomery M.W.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Venta P.J., Li J., Yuzbasiyan-Gurkan V., Brewer G.J., Schall W.D.;
RT "Complete sequence of the structural gene for canine von Willebrand
factor and identification of a mutation causing Scottish terrier von
Willebrand's disease.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 1234-1669 FROM N.A.
RC TISSUE=Blood;
RA Mancuso D.J., Christopherson P.A., Kroner P.A., Montgomery R.R.;
RT "The canine von Willebrand factor gene: sequence and expression of
a region encoding the glycoprotein IB/IX binding domain.";
RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: IMPORTANT IN THE MAINTENANCE OF HOMEOSTASIS, IT
PARTICIPATES IN PLATELET-VESSEL WALL INTERACTIONS BY FORMING A
NONCOVALENT COMPLEX WITH COAGULATION FACTOR VIII AT THE SITE OF
VASCULAR INJURY (BY SIMILARITY).
CC -!- SUBUNIT: MULTIMERIC (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: BLOOD.
CC -!- PTM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR
INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).
CC -!- SIMILARITY: Contains 3 WVEA domains.
CC -!- SIMILARITY: Contains 3 WVEA domains.
CC -!- SIMILARITY: Contains 4 WVEF domains.
CC -!- SIMILARITY: Contains 1 C-terminal cysteine knot-like (CTCK) domain.
CC -!- SIMILARITY: SOME, TO SILKWORM HEMOCYTIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L76227; AAB05549.1; --
DR EMBL; L16903; AAA30903.1; --
DR EMBL; AF099154; AAD04919.1; --
DR EMBL; U66246; AAB93766.1; --
DR HSSP; P04275; 1AUQ.
DR InterPro; IPR006208; Cys_knot.
DR InterPro; IPR006207; Cys_knot_.
DR InterPro; IPR000004; SapB.
DR InterPro; IPR002919; TIL_Cysrich.
DR InterPro; IPR002035; WVF_A.
```


DR InterPro; IPR001007; VWF C.
 DR InterPro; IPR001846; VWF D.
 DR Pfam; PF000007; Cys_knot_1.
 DR Pfam; PF01826; TIL; 4.
 DR Pfam; PF00092; vwa; 3.
 DR Pfam; PF00093; vwc; 3.
 DR Pfam; PF00094; vwd; 4.
 DR PRINTS; PR00453; VWFADOMAIN.
 DR SMART; SM00041; CT; 1.
 DR SMART; SM00327; VWA; 3.
 DR SMART; SM00214; VWC; 5.
 DR SMART; SM00216; VWD; 4.
 DR PROSITE; PS01185; CTCK_1; 1.
 DR PROSITE; PS01225; CTCK_2; 1.
 DR PROSITE; PS02334; VWF_A; 3.
 DR PROSITE; PS01208; VWF_C_1; 3.
 DR PROSITE; PS0184; VWF_C_2; 3.
 DR Blood coagulation; Platelet;
 KW Plasma; Hemostasis; Repeat;
 KW Signal; 1 22
 FT PROPEP 23 763
 FT CHAIN 764 2813
 FT DOMAIN 35 179
 FT DOMAIN 388 541
 FT DOMAIN 764 787
 FT DOMAIN 788 833
 FT DOMAIN 826 853
 FT DOMAIN 867 1013
 FT DOMAIN 1277 1453
 FT DOMAIN 1498 1665
 FT DOMAIN 1691 1871
 FT DOMAIN 1950 2102
 FT DOMAIN 2216 2261
 FT DOMAIN 2255 2328
 FT DOMAIN 2429 2495
 FT DOMAIN 2580 2645
 FT DOMAIN 2724 2812
 FT SITE 531 533
 FT SITE 698 700
 FT SITE 2507 2509
 FT DISULFID 767 808
 FT DISULFID 776 804
 FT DISULFID 810 821
 FT DISULFID 867 936
 FT DISULFID 889 1031
 FT DISULFID 898 993
 FT DISULFID 914 921
 FT DISULFID 1060 1084
 FT DISULFID 1071 1111
 FT DISULFID 1089 1091
 FT DISULFID 1153 1165
 FT DISULFID 1149 1169
 FT DISULFID 1126 1130
 FT DISULFID 1196 1199
 FT DISULFID 1234 1237
 FT DISULFID 1272 1458
 FT DISULFID 1669 1670
 FT DISULFID 1686 1872
 FT DISULFID 1879 1904
 FT DISULFID 1899 1940
 FT DISULFID 1972 2123
 FT DISULFID 1950 2085
 FT DISULFID 1927 2088
 FT DISULFID 1993 2001
 FT DISULFID 2724 2774
 FT DISULFID 2739 2788
 FT DISULFID 2750 2804
 FT DISULFID 2754 2806
 FT DISULFID ? 2811
 FT CARBOHYD 99 99
 FT CARBOHYD 156 156
 FT CARBOHYD 211 211
 FT CARBOHYD. 666 666

FT CARBOHYD 857 857
 FT CARBOHYD 1231 1231
 FT CARBOHYD 1515 1515
 FT CARBOHYD 1574 1574
 FT CARBOHYD 2223 2223
 FT CARBOHYD 2290 2290
 FT CARBOHYD 2357 2357
 FT CARBOHYD 2400 2400
 FT CARBOHYD 2546 2546
 FT CARBOHYD 2585 2585
 FT CARBOHYD 2790 2790
 FT CONFLICT 55 55
 FT CONFLICT 70 70
 FT CONFLICT 266 266
 FT CONFLICT 280 280
 FT CONFLICT 409 411
 FT CONFLICT 994 994
 FT CONFLICT 1021 1021
 FT CONFLICT 2381 2381
 FT CONFLICT 2406 2406
 FT CONFLICT 2813 2813
 SQ SEQUENCE 2813 AA; 309716 MW; 5DF93E1E5E72F80C CRC64;
 Query Match 26.5%; Score 103.5; DB 1; Length 2813;
 Best Local Similarity 28.8%; Pred. No. 0.002;
 Matches 19; Conservative 15; Mismatches 27; Indels 5; Gaps 2;
 QY 2 GFGGLGGRGKPSNEIFSRCDGRQRCFQPNV-VKPLCIKICAGCVCRILYLNKKVC 60
 DB 646 GFCAL----SCPGQVYLQCGTPCNMTCSRSLSPEDCNVCLGCFPGGLYLDRGDC 701
 QY 61 VPRSKC 66
 DB 702 VPKAQC 707
 RESULT 6
 VWF_BOVIN STANDARD; PRT; 937 AA.
 AC P80012; Q28011;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Von Willebrand factor precursor (VWF) (Fragment).
 GN VWF OR F8VWF
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97307751; PubMed=9165093;
 RA Janel N., Ribba A.S.; Cherel G.; Kerbirou-Nabias D., Meyer D.;
 RT "Primary structure of the propeptide and factor VIII-binding domain of
 RT bovine von Willebrand factor";
 RL Biochim. Biophys. Acta 1339:4-8(1997).
 RN [2]
 RP SEQUENCE OF 1-177 FROM N.A.
 RX MEDLINE=96144290; PubMed=8566794;
 RA Janel N., Schwachtgen J.L., Bakhshi M.R., Barek L., Meyer D.,
 RA Kerbirou-Nabias D.;
 RT "Comparison of the 5'-flanking sequences of the human and bovine von
 RT Willebrand factor-encoding genes reveals alternation of highly
 RT homologous domains with species-specific Alu-type repeats.";
 RL Gene 167:291-295(1995).
 RN [3]
 RP PARTIAL SEQUENCE.
 RX MEDLINE=91192039; PubMed=1707363;
 RA Fujisawa I., Takiguchi J., Sekiya F., Goto A., Miake F., Saito Y.;
 RT "Monoclonal antibodies that inhibit binding of propolypeptide of von
 RT Willebrand factor to collagen. Localization of epitopes";
 RL Eur. J. Biochem. 196:673-677(1991).
 RN [4]

PARTIAL SEQUENCE.
RX MEDLINE=96028118; PubMed=7588715;
RA Takagi J., Aoyama T., Ueki S., Ohba H., Saito Y., Lorand L.;
RT "Identification of factor-XIIa-reactive glutamyl residues in the
RL propolypeptide of bovine von Willebrand factor.";
RL Eur. J. Biochem. 232:773-777(1995).
CC -!- FUNCTION: IMPORTANT IN THE MAINTENANCE OF HEMOSTASIS, IT
CC PARTICIPATES IN PLATELET-VESSEL WALL INTERACTIONS BY FORMING A
CC NONCOVALENT COMPLEX WITH COAGULATION FACTOR VIII AT THE SITE OF
CC VASCULAR INJURY.
CC -!- SUBUNIT: MULTIMERIC (BY SIMILARITY).
CC -!- FM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR
CC INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).
CC -!- SIMILARITY: Contains at least 3 VWFD domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Y09353; CAA70525.1; -;
DR EMBL; U28147; AAA96953.1; -;
DR InterPro; IPR006207; Cys_knot C.
DR InterPro; IPR002919; TIL_Cysrich.
DR InterPro; IPR001007; VWF_C.
DR InterPro; IPR001846; VWF_D.
DR Pfam; PF01826; TIL; 2.
DR Pfam; PF00094; vwd; 3.
DR SMART; SM00214; VWD; 1.
DR SMART; SM00216; VWD; 2.
DR PROSITE; PS01185; CTCK_1; PARTIAL.
DR PROSITE; PS01225; CTCK_2; PARTIAL.
DR PROSITE; PS01208; VWFC_1; PARTIAL.
DR Blood coagulation; Platelet; Glycoprotein; Extracellular matrix;
KW Plasma; Hemostasis; Cell adhesion; Repeat; Signal.
FT SIGNAL 1 22
FT PROPEP 23 762
FT CHAIN 763 >937
FT DOMAIN 35 179
FT DOMAIN 387 540
FT DOMAIN 763 786
FT DOMAIN 787 832
FT DOMAIN 825 852
FT DOMAIN 841 >937
FT SITE 410 410
FT SITE 414 414
FT SITE 605 605
FT DISULFID 766 807
FT DISULFID 775 803
FT DISULFID 913 920
FT CARBOHYD 99 99
FT CARBOHYD 156 156
FT CARBOHYD 211 211
FT CARBOHYD 665 665
FT CARBOHYD 856 856
FT CONFLICT 330 330
FT CONFLICT 523 523
FT CONFLICT 528 528
FT NON TER 937 937
SQ SEQUENCE 937 AA; 102599 MW; 9BFC94A254A5629 CRC64;

Query Match 26.0%; Score 101.5; DB 1; Length 937;
Best Local Similarity 28.6%; Pred. No. 0.0013;
Matches 16; Conservative 13; Mismatches 26; Indels 1; Gaps 1;

Qy 12 CPSNEISRCDCRCPCNVV-PKPLCIKICAGCVCRGLYLNKKVCVPRSKC 66
Dd 651 CPHGVVQQCGTGNLTCRSLSHPDDECTEVCLEGCFCPPGLFLDGTSCVPRQA 706

RESULT 7
ZAN_RABIT
ID_ZAN_RABIT STANDARD; PRT; 2282 AA.
AC P57399;
DT 16-OCT-2001 (Rel. 40, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Zonadhesin (Fragment).
GN ZAN.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Theria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=21573741; PubMed=11717130;
RA Lea I.A., Sivashanmugam P., O'Rand M.G.;
RT "Zonadhesin: characterization, localization, and zona pellucida
RT binding.";
RL Biol. Reprod. 65:1691-1700(2001).
CC -!- FUNCTION: BINDS IN A SPECIES-SPECIFIC MANNER TO THE ZONA PELLUCIDA
CC OF THE EGG. MAY BE INVOLVED IN GAMETE RECOGNITION AND/OR
CC SIGNALING (BY SIMILARITY).
CC -!- SUBUNIT: PROBABLY FORMS COVALENT OLIGOMERS (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, EXCLUSIVELY ON THE
CC APICAL REGION OF THE SPERM HEAD (BY SIMILARITY)
CC -!- DOMAIN: THE MAM DOMAINS PROBABLY MEDIATE SPERM ADHESION TO THE
CC ZONA PELLUCIDA.
CC -!- DOMAIN: DURING SPERM MIGRATION THROUGH THE REPRODUCTIVE TRACTS,
CC THE MUCIN-LIKE DOMAIN MIGHT INHIBIT INAPPROPRIATE TRAPPING OF
CC SPERMATOZOA OR PROMOTING ADHESION TO THE OVIDUCTAL ISTHMUS.
CC -!- DOMAIN: THE VWFD DOMAINS 2 AND 3 MAY MEDIATE COVALENT
CC OLIGOMERIZATION (BY SIMILARITY TO HUMAN INTESTINAL MUCIN MUC2).
CC -!- SIMILARITY: Contains 5 VWFD domains.
CC -!- SIMILARITY: Contains 1 EGF-like domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF244982; AAF63342.2; -;
DR HSSP; P56682; ICCV.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR000998; MAM_domain.
DR InterPro; IPR002919; TIL_Cysrich.
DR InterPro; IPR003328; TILa_Cysrich.
DR InterPro; IPR001007; VWF_C.
DR InterPro; IPR001846; VWF_D.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF00629; MAM; 2.
DR Pfam; PF01826; TIL; 5.
DR Pfam; PF02345; TILa; 5.
DR Pfam; PF00094; vwd; 4.
DR SMART; SM00137; MAM; 1.
DR SMART; SM00214; VWC; 2.
DR SMART; SM00216; VWD; 4.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 4.
DR PROSITE; PS00740; MAM_1; FALSE_NEG.
DR PROSITE; PS00060; MAM_2; 2.
KW Glycoprotein; Transmembrane; EGF-like domain; Cell adhesion;
Repeat.
KW NON TER 1 1
FT DOMAIN <1 2235 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 2236 2256 POTENTIAL.
FT DOMAIN 2257 2282 CYTOPLASMIC (POTENTIAL).


```

FT CHAIN 25 115 ACCESSORY GLAND PROTEIN ACP62F.
FT DOMAIN 34 88 TIL.
SQ SEQUENCE 115 AA; 12570 MW; 4326AA6F6C32291D CRC64;

Query Match 25.1%; Score 98; DB 1; Length 115;
Best Local Similarity 34.4%; Pred. No. 0.00053;
Matches 22; Conservative 9; Mismatches 31; Indels 2; Gaps 2;

QY 4 GGLGGKQPSNIFGRCDRCQRCFCNVVVPKPLCIKICAPGCVCLRGYLRKK-KVCVP 62
Db 26 GWGPKVDCITANGTQTECPVACPETCEYSGNGP-CVVMGAPCVCKPGYVNERIPACVL 84
QY 63 RSKC 66
Db 85 RSDC 88

RESULT 9
VWF_HUMAN
ID _VWF_HUMAN STANDARD; PRT; 2813 AA.
AC P04275;
DT 20-MAR-1987 (Rel. 04, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Von Willebrand factor precursor (vWF).
GN VWF OR F8VWF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90062044; PubMed=2584182;
RA Mancuso D.J., Tuley E.A., Westfield L.A., Worrall N.K.,
RA Shelton-Inloes B.B., Sorace J.M., Alevy Y.G., Sadler J.E.;
RT "Structure of the gene for human von Willebrand factor.";
RL J. Biol. Chem. 264:19514-19527(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=87016349; PubMed=3489923;
RA Bonthron D., Orr E.C., Mitsock L.M., Ginsburg D., Handin R.I.,
RA Orkin S.H.;
RT "Nucleotide sequence of pre-pro-von Willebrand factor cDNA.";
RL Nucleic Acids Res. 14:7125-7128(1986).
RN [3]
RP SEQUENCE OF 1-120 FROM N.A., AND SEQUENCE OF 23-56.
RC TISSUE=Umbilical vein endothelial cells;
RX MEDLINE=87213253; PubMed=3495266;
RA Shelton-Inloes B.B., Broze G.J. Jr., Miletich J.P., Sadler J.E.;
RT "Evolution of human von Willebrand factor: cDNA sequence
RT polymorphisms, repeated domains, and relationship to von Willebrand
RT antigen II.";
RL Biochem. Biophys. Res. Commun. 144:657-665(1987).
RN [4]
RP SEQUENCE OF 1-1400 FROM N.A.
RX MEDLINE=87004550; PubMed=3019665;
RA Verweij C.L., Diergaarde P.J., Hart M., Pannekoek H.;
RT "Full-length von Willebrand factor (vWF) cDNA encodes a highly
RT repetitive protein considerably larger than the mature vWF subunit.";
RL EMBO J. 5:1839-1847(1986).
RN [5]
RP ERRATUM.
RA Verweij C.L., Diergaarde P.J., Hart M., Pannekoek H.;
RL EMBO J. 5:3074-3074(1986).
RN [6]
RP SEQUENCE OF 764-2813.
RX MEDLINE=86269895; PubMed=3524673;
RA Titani K., Kumar S., Takio K., Ericsson L.H., Wade R.D., Ashida K.,
RA Walsh K.A., Chopek M.W., Sadler J.E., Fujikawa K.;
RT "Amino acid sequence of human von Willebrand factor.";
RL Biochemistry 25:3171-3184(1986).
RN [7]
RP SEQUENCE OF 781-1424 FROM N.A.

```

```

RX MEDLINE=86269894; PubMed=3488076;
RA Shelton-Inloes B.B., Titani K., Sadler J.E.;
RT "cDNA sequences for human von Willebrand factor reveal five types of
RT repeated domains and five possible protein sequence polymorphisms.";
RL Biochemistry 25:3164-3171(1986).
RN [8]
RP SEQUENCE OF 764-873 AND 1289-2813 FROM N.A.
RX MEDLINE=86016708; PubMed=2864688;
RA Sadler J.E., Shelton-Inloes B.B., Sorace J.M., Harlan J.M.,
RA Titani K., Davie E.W.;
RT "Cloning and characterization of two cDNAs coding for human von
RT Willebrand factor.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:6394-6398(1985).
RN [9]
RP SEQUENCE OF 990-1947 FROM N.A.
RX MEDLINE=91105089; PubMed=1988024;
RA Mancuso D.J., Tuley E.A., Westfield L.A., Lester-Mancuso T.L.,
RA Le Beau M.M., Sorace J.M., Sadler J.E.;
RT "Human von Willebrand factor gene and pseudogene: structural analysis
RT and differentiation by polymerase chain reaction.";
RL Biochemistry 30:253-269(1991).
RN [10]
RP SEQUENCE OF 2731-2813 FROM N.A.
RX MEDLINE=85269603; PubMed=3875078;
RA Verweij C.L., de Vries C.J.M., Distel B., van Zonneveld A.-J.,
RA Geurts van Kessel A., van Mourik J.A., Pannekoek H.;
RT "Construction of cDNA coding for human von Willebrand factor using
RT antibody probes for colony-screening and mapping of the chromosomal
RT gene.";
RL Nucleic Acids Res. 13:4699-4717(1985).
RN [11]
RP SEQUENCE OF 1-177 FROM N.A.
RX MEDLINE=88111704; PubMed=2828057;
RA Bonthron D., Orkin S.H.;
RT "The human von Willebrand factor gene. Structure of the 5' region.";
RL Eur. J. Biochem. 171:51-57(1988).
RN [12]
RP SEQUENCE OF 2621-2813 FROM N.A.
RX MEDLINE=85244588; PubMed=3874428;
RA Ginsburg D., Handin R.I., Bonthron D.T., Donlon T.A., Bruns G.A.P.,
RA Latt S.A., Orkin S.H.;
RT "Human von Willebrand factor (vWF): isolation of complementary DNA
RT (cDNA) clones and chromosomal localization.";
RL Science 228:1401-1406(1985).
RN [13]
RP SEQUENCE OF 2731-2813 FROM N.A.
RX MEDLINE=85201687; PubMed=3873280;
RA Lynch D.C., Zimmerman T.S., Collins C.J., Brown M., Morin M.J.,
RA Ling E.H., Livingston D.M.;
RT "Molecular cloning of cDNA for human von Willebrand factor:
RT authentication by a new method.";
RL Cell 41:49-56(1985).
RN [14]
RP REVISIONS.
RA Lynch D.C.;
RL Submitted (JUL-1991) to the EMBL/GenBank/DBJ databases.
RN [15]
RP SEQUENCE OF 2731-2813 FROM N.A.
RX MEDLINE=87260814; PubMed=3496594;
RA Collins C.J., Underdahl J.P., Levene R.B., Ravera C.P.,
RA Morin M.J., Dombalagian M.J., Ricca G., Livingston D.M.,
RA Lynch D.C.;
RT "Molecular cloning of the human gene for von Willebrand factor and
RT identification of the transcription initiation site.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:4393-4397(1987).
RN [16]
RP DISULFIDE BONDS.
RX MEDLINE=88163465; PubMed=3502076;
RA Marti T., Rosselet S.J., Titani K., Walsh K.A.;
RT "Identification of disulfide-bridged substructures within human von
RT Willebrand factor.";
RL Biochemistry 26:8099-8109(1987).
RN [17]

```

RP STRUCTURE OF CARBOHYDRATES.
RX MEDLINE=86274702; PubMed=3089784;
RA Samor B., Michalski J.C., Debray H., Mazurier C., Goudemand M.,
van Halbeek H., Vliegthart J.F.G., Montreuil J.,
RT "Primary structure of a new tetraantennary glycan of the N-
acetylglucosaminic type isolated from human factor VIII/von
Willebrand factor.";
RL Eur. J. Biochem. 158:295-298(1986).
RN [18]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 1261-1468.
RX MEDLINE=98221174; PubMed=9553097;
RA Emsley J., Cruz M., Handin R., Liddington R.;
RT "Crystal structure of the von Willebrand factor A1 domain and
implications for the binding of platelet glycoprotein Ib.";
RL J. Biol. Chem. 273:10396-10401(1998).
RN [19]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 1685-1873.
RX MEDLINE=97472999; PubMed=9331419;
RA Huizinga E.G., Martijn van der Plas R., Kroon J., Sixma J.J., Gros P.;
RT "Crystal structure of the A3 domain of human von Willebrand factor:
implications for collagen binding.";
RL Structure 5:1147-1156(1997).
RN [20]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1686-1872.
RX MEDLINE=97460108; PubMed=9312128;
RA Bienkowska J., Cruz M., Atiemo A., Handin R., Liddington R.;
RT "The von Willebrand factor A3 domain does not contain a metal ion-
dependent adhesion site motif.";
RL J. Biol. Chem. 272:25162-25167(1997).
RN [21]
RP VARIANTS TRP-1597 AND ASP-1607.
RX MEDLINE=89264495; PubMed=2786201;
RA Ginsburg D., Konkle B.A., Gill J.C., Montgomery R.R.,
Bockenstedt P.L., Johnson T.A., Yang A.Y.;
RT "Molecular basis of human von Willebrand disease: analysis of
platelet von Willebrand factor mRNA.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:3723-3727(1989).
RN [22]
RP VARIANT THR-1628.
RX MEDLINE=91196734; PubMed=1673047;
RA Iannuzzi M.C., Hidaka N., Boehnke M., Bruck M.E., Hanna W.T.,
Collins F.S., Ginsburg D.;
RT "Analysis of the relationship of von Willebrand disease (vWD) and
hereditary hemorrhagic telangiectasia and identification of a
potential type IIA vWD mutation (Ile865 to Thr).";
RL Am. J. Hum. Genet. 48:757-763(1991).
RN [23]
RP VARIANTS NORMANDY-2 TRP-816 AND NORMANDY-3 GLN-854.
RX MEDLINE=92001464; PubMed=1832934;
RA Gaucher C., Mercier B., Jorieux S., Oufkir D., Mazurier C.;
RT "Identification of two point mutations in the von Willebrand factor
gene of three families with the 'Normandy' variant of von Willebrand
disease.";
RL Br. J. Haematol. 78:506-514(1991).
RN [24]
RP VARIANT CYS-1308.
RX MEDLINE=92104315; PubMed=1761120;
RA Donner M., Andersson A.-M., Kristoffersson A.-C., Nilsson I.M.,
Dahlback B., Holmberg L.;
RT "An Arg545->Cys545 substitution mutation of the von Willebrand
factor in type IIB von Willebrand's disease.";
RL Eur. J. Haematol. 47:342-345(1991).
RN [25]
RP VARIANTS TRP-1306; CYS-1308 AND PRO-1613.
RX MEDLINE=91185601; PubMed=2010538;
RA Randi A.M., Rabinowitz I., Mancuso D.J., Mannucci P.M., Sadler J.E.;
RT "Molecular basis of von Willebrand disease type IIB. Candidate
mutations cluster in one disulfide loop between proposed platelet
glycoprotein Ib binding sequences.";
RL J. Clin. Invest. 87:1220-1226(1991).
RN [26]
RP VARIANTS TRP-1306; CYS-1308; MET-1316; GLN-1341 AND HIS-1399.
RX MEDLINE=91185602; PubMed=1672694;

Query Match 25.1%; Score 98; DB 1; Length 2813;
Best Local Similarity 26.0%; Pred. No. 0.008;
Matches 19; Conservative 13; Mismatches 27; Indels 14; Gaps 2;
QY 8 GRG-----KCPSEIFSRCDGRFCQFCFNV-VPKPLCIKICAPGCVCRIGYL 53
DB 635 GRGVAVWRPGRCELNCPKQVYLGQTFNLTCSRSLSPDECNACLEGCFPPGLY 694
QY 54 RNKKKVCVPRSKC 66
DB 695 MDERGDCVPKAQ 707
RESULT 10
CVP6_PIMHY STANDARD; PRT; 77 AA.
AC Q8TOW0;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Cysteine-rich venom protein 6 precursor.
GN CVP6.
OS Pimpla hypochondriaca (Parasitoid wasp).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Hymenoptera; Apocrita; Ichneumonidae;
OC Ichneumonidae; Pimplinae; Pimpla.
OX NCBI_TaxID=135724;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Venom;
RA Parkinson N.M.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Secreted (potential).
CC -!- SIMILARITY: Contains 1 TIL (Trypsin inhibitory-like) domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ438997; CAD27742.1; -
DR InterPro; IPR002919; TIL_Cysrich.
DR Pfam; PF01826; TIL; 1.
KW Signal.
FT SIGNAL 1 21 Cysteine-rich venom protein 6.
FT CHAIN 22 77 TIL.
FT DOMAIN 24 76 BY SIMILARITY.
FT DISULFID 24 56 BY SIMILARITY.
FT DISULFID 33 52 BY SIMILARITY.
FT DISULFID 37 48 BY SIMILARITY.
FT DISULFID 41 76 BY SIMILARITY.
FT DISULFID 58 70 BY SIMILARITY.
SQ SEQUENCE 77 AA; 8184 MW; 6889CB5F40D63DD6 CRC64;
Query Match 24.8%; Score 97; DB 1; Length 77;
Best Local Similarity 32.7%; Pred. No. 0.0048;
Matches 18; Conservative 6; Mismatches 29; Indels 2; Gaps 1;
QY 12 CPSNEIFSRCDGRFCFNVVPPKLCIKICAPGCVCRIGVIRNKKKVCVPRSKC 66
DB 24 CGPNRVYKSCGTGCPETCEN--PDPCDCHQGCFCCKGLLDGNCISPDKC 76
RESULT 11
VWF_PIG STANDARD; PRT; 2482 AA.
ID -VWF_PIG
AC Q28833;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)

28-FEB-2003 (Rel. 41, Last annotation update)
Von Willebrand factor precursor (vWF) (Fragment).
vWF OR F8VWF.
Sus scrofa (Pig).
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
NCBI_TaxID=9823;
[1]
SEQUENCE FROM N.A.
Seaman W.T., Read M.S., Bellinger D.A., Nichols T.C.;
Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE OF 397-553 FROM N.A.
MEDLINE=93356762; PubMed=8352759;
Laveigne J.M., Piao Y.C., Ferreira V., Kerbiriou-Nabias D.,
Bahnak B.R., Meyer D.;
"Primary structure of the factor VIII binding domain of human, porcine
and rabbit von Willebrand factor.";
Biochem. Biophys. Res. Commun. 194:1019-1024(1993).
CC -!- FUNCTION: IMPORTANT IN THE MAINTENANCE OF HOMEOSTASIS, IT
PARTICIPATES IN PLATELET-VESSEL WALL INTERACTIONS BY FORMING A
NONCOVALENT COMPLEX WITH COAGULATION FACTOR VIII AT THE SITE OF
VASCULAR INJURY (BY SIMILARITY).
CC -!- SUBUNIT: MULTIMERIC (BY SIMILARITY).
CC -!- PTM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR
INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).
CC -!- SIMILARITY: Contains 3 vWF domains.
CC -!- SIMILARITY: Contains 3 vWFC domains.
CC -!- SIMILARITY: Contains 1 C-terminal cysteine knot-like (CTCK) domain.
CC -!- SIMILARITY: SOME, TO SILKWORM HEMOCYTIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; AF052036; AAC06229.1; -.
DR EMBL; S64541; AAB27829.2; -.
DR PIR; P0563; PNO563.
DR HSSP; P04275; LATZ.
DR InterPro; IPR006208; Cys_knot.
DR InterPro; IPR006207; Cys_knot_C.
DR InterPro; IPR002919; TIL_Cysrich.
DR InterPro; IPR002035; vWF_A.
DR InterPro; IPR001007; vWF_C.
DR InterPro; IPR001846; vWF_D.
DR Pfam; PF00007; Cys_knot; 1.
DR Pfam; PF01826; TIL; 3.
DR Pfam; PF00092; vwa; 3.
DR Pfam; PF00093; vwc; 3.
DR Pfam; PF00094; vwd; 3.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00041; CT; 1.
DR SMART; SM00327; VWA; 3.
DR SMART; SM00214; VWC; 5.
DR SMART; SM00216; VWD; 3.
DR PROSITE; PS01185; CTCK 1; 1.
DR PROSITE; PS01225; CTCK 2; 1.
DR PROSITE; PS02234; VWF_A; 3.
DR PROSITE; PS01208; VWF_C; 1; 3.
DR PROSITE; PS0184; VWF_D; 3.
KW Blood coagulation; Platelet; Glycoprotein; Extracellular matrix;
KW Plasma; Hemostasis; Repeat; Cell adhesion.
FT NON_TER 1 1
FT PROPEP <1 437 BY SIMILARITY.
FT CHAIN 438 2482 VON WILLEBRAND FACTOR.
FT DOMAIN 62 215 WFD 2.
FT DOMAIN 438 461 AMINO-TERMINAL.
FT DOMAIN 462 507 E1.
FT DOMAIN 500 527 CX.
DOMAIN 541 687
FT DOMAIN 947 1127 VWF 3.
FT DOMAIN 1167 1334 VWF 1.
FT DOMAIN 1360 1540 VWF 2.
FT DOMAIN 1619 1771 VWF 4.
FT DOMAIN 1885 1930 E2.
FT DOMAIN 1924 1997 VWF 1.
FT DOMAIN 2098 2164 VWF 2.
FT DOMAIN 2249 2314 VWF 3.
FT DOMAIN 2393 2481 CTCK.
SITE 2176 2178
FT DISULFID 441 482
FT DISULFID 450 478
FT DISULFID 484 495
FT DISULFID 541 670
FT DISULFID 563 705
FT DISULFID 572 667
FT DISULFID 588 595
FT DISULFID 734 758
FT DISULFID 763 765
FT DISULFID 827 839
FT DISULFID 823 843
FT DISULFID 800 804
FT DISULFID 870 873
FT DISULFID 908 911
FT DISULFID 942 1128
FT DISULFID 1338 1339
FT DISULFID 1355 1541
FT DISULFID 1548 1573
FT DISULFID 1568 1609
FT DISULFID 1641 1792
FT DISULFID 1619 1754
FT DISULFID 1596 1757
FT DISULFID 1662 1670
FT DISULFID 2393 2443
FT DISULFID 2408 2457
FT DISULFID 2419 2473
FT DISULFID 2423 2475
FT DISULFID ? 2480
FT CARBOHYD 340 340
FT CARBOHYD 905 905
FT CARBOHYD 1184 1184
FT CARBOHYD 1243 1243
FT CARBOHYD 1640 1640
FT CARBOHYD 1892 1892
FT CARBOHYD 1959 1959
FT CARBOHYD 2026 2026
FT CARBOHYD 2069 2069
FT CARBOHYD 2215 2215
FT CARBOHYD 2254 2254
FT CARBOHYD 2459 2459
SQ SEQUENCE 2482 AA; 272394 MW; D499B7DDFBBCAEDD CRC64;
Query Match 23.4%; Score 91.5; DB 1; Length 2482;
Best Local Similarity 28.8%; Pred. No. 0.036;
Matches 19; Conservative 11; Mismatches 31; Indels 5; Gaps 2;
QY 2 GFGLGGRGKCPSEIFRSCDGRQCFQPNV-VPEKLCIKICAPGCVCRILNKKKVC 60
Db 320 GFCAI-SCPPGQVYVLAQGTGTPCNLCRSLSYFDECAEDCUEGCFPPGLVDGSGDC 375
QY 61 VPRSKC 66
Db 376 VPRAQC 381
RESULT 12
HMCT_BOMMO
ID HMCT_BOMMO STANDARD; PRT; 3133 AA.
AC P98092;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Hemocytin precursor (Humoral lectin).
OS Bombyx mori (Silk moth).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Bombycoidea;
OC Bombycidae; Bombyx.
OX NCBI_TaxID=7091;
RN [1]_TaxID=7091;
RP SEQUENCE FROM N.A.
RC STRAIN=Fuyou X Tokai; TISSUE=Hemocyte;
RX MEDLINE=95178544; PubMed=7873598;
RA Kotani E., Yamakawa M., Iwamoto S.-I., Tashiro M., Mori H., Sumida M.,
RA Matsubara F., Tanikawa K., Kadohira-Okuda K., Kato Y., Mori H.,
RT "Cloning and expression of the gene of hemocytin, an insect humoral
RT lectin which is homologous with the mammalian von Willebrand
RT factor.";
RL Biochim. Biophys. Acta 1260:245-258(1995).
RN [2]
RP SEQUENCE OF 2221-3133 FROM N.A.
RA Kotani E., Iwamoto S.-I., Tashiro M., Mori H., Sumida M.,
RA Matsubara F., Tanikawa K., Kadohira-Okuda K., Kato Y., Mori H.,
RL Submitted (JAN-1993) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ADHESIVE PROTEIN AND RELATES TO HEMOSTASIS OR
CC ENCAPSULATION OF FOREIGN SUBSTANCES FOR SELF-DEFENSE.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED IN HEMOCYTES DURING LARVAL-PUPAL
CC METAMORPHOSIS.
CC -!- INDUCTION: HEMAGGLUTINATION ACTIVITY IS INCREASED BY BACTERIAL
CC OR VIRAL INFECTION AND INHIBITED BY D-MANNOSE, N-ACETYL-D-
CC GALACTOSAMINE AND D-MALTOSE.
CC -!- PTM: MAY BE CONVERTED INTO THE 260 kDa MATURE HEMOCYTIN BY
CC PROTEOLYSIS.
CC -!- SIMILARITY: TO MAMMALIAN VON WILLEBRAND FACTOR.
CC -!- SIMILARITY: THE N-TERMINAL DOMAIN SHOWS SOME SIMILARITY TO THAT
CC OF HUMAN MUCIN 2.
CC -!- SIMILARITY: Contains 2 F5/8 type C domains.
CC -!- SIMILARITY: Contains 2 VF5C domains.
CC -!- SIMILARITY: Contains 1 C-terminal cysteine knot-like (CTCK) domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D29738; BAA06160.1; -;
CC EMBL; D14035; BAA03124.1; -;
CC PIR; S52093; S52093.
CC DR HSP: P12259; 1CZT.
DR InterPro; IPR006207; Cys_knot_C.
DR InterPro; IPR000421; FA58_C.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR002919; TIL_Cysrich.
DR InterPro; IPR001007; VWF_C.
DR InterPro; IPR001846; VWF_D.
DR Pfam; PF00754; F5_P8_type_C; 2.
DR Pfam; PF01826; TIL; 6.
DR Pfam; PF00094; vwd; 3.
DR SMART; SM00041; CT; 1.
DR SMART; SM00231; FA58C; 2.
DR SMART; SM00192; LDLA; 1.
DR SMART; SM00216; VWD; 3.
DR PROSITE; PS01185; CTCK_1; 1.
DR PROSITE; PS01225; CTCK_2; 1.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 2.
DR PROSITE; PS00022; FA58C_3; 2.
DR PROSITE; PS01208; VWF_1; FALSE_NEG.
KW Lectin; Glycoprotein; Signal; Repeat; Cell adhesion.
FT SIGNAL 1 ?
FT CHAIN ? 3133
FT DOMAIN 29 131
FT D'.

FT DOMAIN 153 240
FT DOMAIN 248 613
FT DOMAIN 940 1095
FT DOMAIN 1116 1234
FT DOMAIN 1283 1356
FT DOMAIN 1620 1951
FT DOMAIN 1952 2315
FT DOMAIN 2321 2321
FT DOMAIN 2330 2361
FT DOMAIN 2435 2469
FT DOMAIN 2553 2622
FT DOMAIN 2842 2907
FT DOMAIN 2971 3076
FT DOMAIN 895 914
FT DOMAIN 1257 1270
FT DOMAIN 1425 1428
FT DOMAIN 1447 1450
FT DOMAIN 1474 1479
FT DOMAIN 2148 2153
FT DOMAIN 2156 2159
FT DOMAIN 2341 2344
FT DISULFID 940 1095
FT DISULFID 1116 1234
FT DISULFID 2981 3040
FT DISULFID 2991 3054
FT DISULFID 3004 3070
FT DISULFID 3020 3072
FT DISULFID ? 3075
FT CARBOHYD 151 151
FT CARBOHYD 237 237
FT CARBOHYD 564 564
FT CARBOHYD 1170 1170
FT CARBOHYD 1387 1387
FT CARBOHYD 1622 1622
FT CARBOHYD 1727 1727
FT CARBOHYD 1847 1847
FT CARBOHYD 1975 1975
FT CARBOHYD 1985 1985
FT CARBOHYD 2093 2093
FT CARBOHYD 2113 2113
FT CARBOHYD 2161 2161
FT CARBOHYD 2276 2276
FT CARBOHYD 2451 2451
FT CARBOHYD 2647 2647
FT CARBOHYD 2654 2654
FT CARBOHYD 2663 2663
FT CARBOHYD 2794 2794
FT CARBOHYD 2810 2810
FT CARBOHYD 2865 2865
FT CARBOHYD 2929 2929
FT CARBOHYD 2964 2964
FT CARBOHYD 3028 3028
FT VARIANT 1288 1288
FT VARIANT 1305 1305
SQ SEQUENCE 3133 AA; 343350 MW; B5210D5D14A7B2B2 CRC64;
Query Match 23.1%; Score 90.5; DB 1; Length 3133;
Best Local Similarity 27.5%; Pred. No. 0.056;
Matches 19; Conservative 11; Mismatches 26; Indels 13; Gaps 3;
QY 11 KCPSEIFSRCDGRCQFCFN----VVPKPLCI-KICAPGCV-----CRLGYLRNKK 57
DB 769 RCPPEGVYQACAYKCDRLCDHFKKTLIAKGRCLSEMCVDCGVDESVAISNGCGSSRWDE 828
QY 58 KVCVPRSKC 66
DB 829 RTCVPVKDC 837
RESULT 13
BAR3_CHITE
ID BAR3_CHITE
AC Q03376; PRT; 1700 AA.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2003, 08:06:07 ; Search time 35 Seconds
(without alignments)
493.986 Million cell updates/sec

Title: US-09-506-978-1

Perfect score: 391

Sequence: 1 GGFGGLGGRGKCPSEIFSR.....CRGLYLNNKKVCVPSKCG 67

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriaph:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	133	34.0	245	5 Q19964	Q19964 caenorhabdi
2	118	30.2	432	5 Q814B8	Q814B8 caenorhabdi
3	118	30.2	537	5 Q9UI76	Q9UI76 caenorhabdi
4	118	30.2	731	5 Q814B9	Q814B9 caenorhabdi
5	118	30.2	796	5 Q9UI75	Q9UI75 caenorhabdi
6	116.5	29.8	2813	11 Q8CIZ8	Q8CIZ8 mus musculu
7	113	28.9	2155	11 Q08523	Q08523 mus musculu
8	113	28.9	5374	11 Q99ND0	Q99ND0 mus musculu
9	111.5	28.5	166	5 O16488	O16488 caenorhabdi
10	111	28.4	4998	11 Q8CG65	Q8CG65 mus musculu
11	110	28.1	69	5 Q8MX86	Q8MX86 glossina mo
12	109.5	28.0	949	5 P90356	P90356 caenorhabdi
13	108	27.6	84	5 Q81CQ5	Q81CQ5 anopheles s
14	108	27.6	3843	5 Q9VU94	Q9VU94 drosophila
15	108	27.6	3843	5 Q9U5D0	Q9U5D0 drosophila
16	107	27.4	2155	4 Q75443	Q75443 homo sapien

ALIGNMENTS

RESULT 1

Q19964
ID Q19964 PRELIMINARY; PRT; 245 AA.
AC Q19964;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE F32D8.3 protein.
GN F32D8.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Wilkinson J.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z74031; CAA984455.1; -.
DR HSSP; P56682; 1CCV.
DR WormPep; F32D8.3; CE05782.
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR002919; TIL_Cysrich.
DR Pfam; PF01826; TIL; 1.
DR PROSITE; PS01186; EGF 2; 1.
SQ SEQUENCE 245 AA; 27785 MW; 781AE77F9FE784C CRC64;

Query Match 34.0%; Score 133; DB 5; Length 245;

Best Local Similarity 43.1%; Pred. No. 5.9e-09;

Matches 28; Conservative 7; Mismatches 20; Indels 10; Gaps 4;

QY 6 LGGRGKCPSEIFSRGDGRFCNVVPEKPLCIKICAPGCVCLGVLNNKKKVCV----61

```
Db 57 INGENCFMFQSFSGH--ACESTCNN--PDVYCSK-CBPGCTCRNGFVRNSLKLVLPEE 111
Qy 62 -PRSK 65
Db 112 CPRTK 116

RESULT 2
Q814B8
ID Q814B8 PRELIMINARY; PRT; 432 AA.
AC Q814B8;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein Y69H2.3d.
GN Y69H2.3D.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA McMurray A.A.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z98877; CAD56617.1; -.
KW Hypothetical protein.
SQ SEQUENCE 432 AA; 46710 MW; F4C99F0221709D67 CRC64;

Query Match 30.2%; Score 118; DB 5; Length 432;
Best Local Similarity 36.8%; Pred. No. 8.1e-07;
Matches 21; Conservative 11; Mismatches 23; Indels 2; Gaps 2;

Qy 12 CPSNEIFSRCDGRC-QRFPCNV-VPKPLCIKICAPGCVCLGKYLNRKKKVCVPRSKC 66
Db 99 CPVNEVSNECHNPCTEKKCPQKNAPQVNCMLACQVGCSCMDGFRNNQGVCKEAE 155

RESULT 3
Q9U1T6
ID Q9U1T6 PRELIMINARY; PRT; 537 AA.
AC Q9U1T6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein Y69H2.3a.
GN Y69H2.3A.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA McMurray A.A.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z98877; CAB54472.2; -.
KW Hypothetical protein.
SQ SEQUENCE 537 AA; 58089 MW; 96F837138278D168 CRC64;

Query Match 30.2%; Score 118; DB 5; Length 537;
```

```
Best Local Similarity 36.8%; Pred. No. 9.9e-07;
Matches 21; Conservative 11; Mismatches 23; Indels 2; Gaps 2;

Qy 12 CPSNEIFSRCDGRC-QRFPCNV-VPKPLCIKICAPGCVCLGKYLNRKKKVCVPRSKC 66
Db 99 CPVNEVSNECHNPCTEKKCPQKNAPQVNCMLACQVGCSCMDGFRNNQGVCKEAE 155

RESULT 4
Q814B9
ID Q814B9 PRELIMINARY; PRT; 731 AA.
AC Q814B9;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein Y69H2.3c.
GN Y69H2.3C.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA McMurray A.A.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z98877; CAD56616.1; -.
KW Hypothetical protein.
SQ SEQUENCE 731 AA; 77752 MW; FC23E1075646733E CRC64;

Query Match 30.2%; Score 118; DB 5; Length 731;
Best Local Similarity 36.8%; Pred. No. 1.3e-06;
Matches 21; Conservative 11; Mismatches 23; Indels 2; Gaps 2;

Qy 12 CPSNEIFSRCDGRC-QRFPCNV-VPKPLCIKICAPGCVCLGKYLNRKKKVCVPRSKC 66
Db 99 CPVNEVSNECHNPCTEKKCPQKNAPQVNCMLACQVGCSCMDGFRNNQGVCKEAE 155

RESULT 5
Q9U1T5
ID Q9U1T5 PRELIMINARY; PRT; 796 AA.
AC Q9U1T5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein Y69H2.3b.
GN Y69H2.3B.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA McMurray A.A.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z98877; CAB54473.2; -.
KW Hypothetical protein.
SQ SEQUENCE 796 AA; 84881 MW; 5945C60C014623D0 CRC64;
```

```
Query Match      30.2%; Score 118; DB 5; Length 796;
Best Local Similarity 36.8%; Pred. No. 1.4e-06;
Matches 21; Conservative 11; Mismatches 23; Indels 2; Gaps 2;

QY 12 CFSNEIFSCDGR-CRQFCPNV-VKPLCIKICAPGCVCLGLYLRNKKVCVPRSKC 66
DB 99 CPVNEVSNECHNPTCKPQKNAPQVNCIMACQVGCSCMDGFRVNNQGVCVKEABC 155

RESULT 6
Q8CIZ8 PRELIMINARY; PRT; 2813 AA.
AC Q8CIZ8;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Von Willebrand factor.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c;
RA Lanting P.J., Westein E., de Groot P.G., Denis C.V.;
RT "Murine von Willebrand Factor.";
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF539800; AAN0781.1; --
SQ SEQUENCE 2813 AA; 308854 MW; E2478C5710E688B5 CRC64;

Query Match      29.8%; Score 116.5; DB 11; Length 2813;
Best Local Similarity 30.3%; Pred. No. 7e-06;
Matches 20; Conservative 16; Mismatches 25; Indels 5; Gaps 2;

QY 2 GFGGLGGRKPKSNEIFSCDGRCPNV-VKPLCIKICAPGCVCLGLYLRNKKVC 60
DB 646 GFCALG---CPQGQVYLQCGNSCLTCRSLSLPDECESEVLEGVCYPPGLYQDGRGDC 701
QY 61 VPRSKC 66
DB 702 VPKAQC 707

RESULT 7
O08523 PRELIMINARY; PRT; 2155 AA.
AC O08523;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Alpha tectorin.
GN TECTA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CD1; TISSUE=Cochlea;
RX MEDLINE=97236843; PubMed=9079715;
RA Legan P.K., Rau A., Keene J.N., Richardson G.P.;
RT "Tectorins: modular matrix proteins of the inner ear homologous to
components of the sperm-egg adhesion system.";
RL J. Biol. Chem. 272:8791-8801(1997).
DR EMBL; X98805; CAA68138.1; --
DR HSSP; P00736; IAPQ.
DR MGD; MGI:109575; Tecta.
DR InterPro; IPR002086; Aldehyde dehydr.
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR001507; Endoglin/CD105.
DR InterPro; IPR000421; FA58 C.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR003886; Nidogen_ext.
```

```
DR InterPro; IPR002919; TIL Cysrich.
DR InterPro; IPR001007; VWF C.
DR InterPro; IPR001846; VWF_D.
DR Pfam; PF01826; TIL; 3.
DR Pfam; PF00094; vwd; 4.
DR Pfam; PF00100; zona_pellucida; 1.
DR SMART; SM00181; EGF; 1.
DR SMART; SM00539; NIDO; 1.
DR SMART; SM00214; VMC; 1.
DR SMART; SM00216; VWD; 4.
DR SMART; SM00241; ZP; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE; PS01186; EGF 2; 3.
DR PROSITE; PS01286; FA58C 2; 1.
DR PROSITE; PS00682; ZP DOMAIN; 1.
SQ SEQUENCE 2155 AA; 239535 MW; 86C5C09AA392B1C5 CRC64;

Query Match      28.9%; Score 113; DB 11; Length 2155;
Best Local Similarity 38.6%; Pred. No. 1.5e-05;
Matches 22; Conservative 7; Mismatches 26; Indels 2; Gaps 2;

QY 11 KCPSEIFSRCDGRCPNVVPKPLCIKICAPGCVCLGLYLRNKKVCVPRSKG 67
DB 983 ECPENSHFEEC-MTCTETCTETALGPICVDSCEGQCDEGY-RLQGSQCVRSECG 1037

RESULT 8
Q99ND0 PRELIMINARY; PRT; 5374 AA.
AC Q99ND0;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE ZAN (Zonadhesin).
GN ZAN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RX MEDLINE=21138439; PubMed=11239002;
RA Wilson M.D., Riener C., Martindale D.W., Schnupf P., Boright A.P.,
Cheung T.L., Hardy D.M., Schwartz S., Scherer S.W., Tsui L.-C.,
Miller W., Koop B.F.;
RT "Comparative analysis of the gene-dense ACHE/TPR2 region on human
chromosome 7q22 with the orthologous region on mouse chromosome 5.";
RL Nucleic Acids Res. 29:1352-1365(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RA Cheung T.L., Wilson M.D., Koop B.F., Hardy D.M.;
RT "Genomic Basis of Inter- and Intra-species Variation in Zonadhesin
Domain Structure.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 3 MAM DOMAINS.
DR EMBL; AF312033; AAK28824.1; --
DR EMBL; AY046056; AAL04416.1; --
DR MGD; MGI:106656; Zan.
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR003645; POLN.
DR InterPro; IPR000998; MAM domain.
DR InterPro; IPR003328; TILA Cysrich.
DR InterPro; IPR002919; TIL Cysrich.
DR InterPro; IPR001007; VWF C.
DR InterPro; IPR001846; VWF_D.
DR Pfam; PF00629; MAM; 3.
DR Pfam; PF01826; TIL; 25.
DR Pfam; PF02345; Tila; 25.
DR Pfam; PF00094; vwd; 4.
DR SMART; SM00274; POLN; 21.
DR SMART; SM00137; MAM; 3.
```

DR SMART; SM00214; VWC; 17.
DR SMART; SM00216; VWD; 4.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 18.
DR PROSITE; PS00060; MAM_2; 3.
KW Glycoprotein.
SQ SEQUENCE 5374 AA; 579536 MW; 90D2D8CFE5DE24EB CRC64;

Query Match 28.9%; Score 113; DB 11; Length 5374;
Best Local Similarity 37.5%; Pred. No. 3.5e-05;
Matches 24; Conservative 8; Mismatches 22; Indels 10; Gaps 3;

QY 11 KCPNSNEIFSRCDGRQCFPCNVVVKPLC-----IKICAPGCVCRGLGYLRNKKKVCVPR 63
DB 4741 KCPANSLYTHCLPTCLPSCSN--PDGRCEGTSKAPSTCRGCVQCPGYLLN-KDTCVHK 4797

QY 64 SKQC 67
DB 4798 NQCC 4801

RESULT 9
ID O16488 PRELIMINARY; PRT; 166 AA.
AC O16488;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical 18.0 kDa protein.
GN B0238.12.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Blanchard M., Bradshaw H.;
RT "The sequence of C. elegans cosmid B0238.";
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF016450; AAB5590.1; --
DR HSP; P56682; 1CCV.
DR WormPep; B0238.12; CE07702.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR002919; TIL_Cysrich.
DR Pfam; PF01826; TIL; 2.
DR PROSITE; PS01186; EGF_2; 2.
KW Hypothetical protein.
SQ SEQUENCE 166 AA; 17971 MW; DA92F2473442ADBD CRC64;

Query Match 28.5%; Score 111.5; DB 5; Length 166;
Best Local Similarity 36.8%; Pred. No. 2.2e-06;
Matches 25; Conservative 11; Mismatches 25; Indels 7; Gaps 4;

QY 4 GGLGGRGKCP---SNEIFSRCDGRQCFPCNVVVKPLCIKICAPG-CVCRGLGYLRNKK-KK 58
DB 27 GQVGGGRLPCRGRNEEYKTCGTACEPSCN--PNFMCTKQINNVCQRCRGYVRNITR 84

QY 59 VCVPRSKC 66

DB 85 QCVRQAQC 92

RESULT 10
Q8CG65 PRELIMINARY; PRT; 4998 AA.
ID Q8CG65;
AC Q8CG65;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE SCO-spondin.
GN SCO-SPONDIN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR; TISSUE=Brain;
RA Goncalves N., Simon-Chazottes D., Creveaux I., Meinzel A.,
Guenet J.-L., Meinzel R.;
RT "Characterization, spatio-temporal expression and chromosomal
assignment of mouse SCO-spondin.";
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ491857; CAD42654.1; --
SQ SEQUENCE 4998 AA; 535019 MW; DA2ABA8DA47DF225 CRC64;

Query Match 28.4%; Score 111; DB 11; Length 4998;
Best Local Similarity 36.7%; Pred. No. 5.9e-05;
Matches 22; Conservative 10; Mismatches 24; Indels 4; Gaps 2;

QY 11 KCPNSNEIFSRCD-GRQCFPCNVVVKPL---CIKICAPGCVCRGLGYLRNKKKVCVPRSKC 66
DB 4776 ECAPGEIWHQHGKLGPCKECTCPENMNTQAWNSNCTEAQAPGCVQLGYFRSQTGLCVPRDHC 4835

RESULT 11
Q8MX86 PRELIMINARY; PRT; 69 AA.
ID Q8MX86;
AC Q8MX86;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Immune reactive putative protease inhibitor Prinh6 (Fragment).
OS Glossina morsitans morsitans.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Hippoboscidae; Glossinidae; Glossina.
OX NCBI_TaxID=37546;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fat body;
RA Akcey S., Hao Z.;
RT "Immune responsive cDNA, Prinh6, characterized from fat body of
Glossina morsitans morsitans.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF368912; AAM46785.1; --
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR002919; TIL_Cysrich.
DR Pfam; PF01826; TIL; 1.
DR PROSITE; PS01186; EGF_2; 1.
KW Protease.
FT NON_TER 1
SQ SEQUENCE 69 AA; 7593 MW; FFBF296663A6C730 CRC64;

Query Match 28.1%; Score 110; DB 5; Length 69;
Best Local Similarity 40.7%; Pred. No. 1.6e-06;
Matches 24; Conservative 5; Mismatches 28; Indels 2; Gaps 2;

QY 8 GRGKCPNSNEIFSRCDGRQCFPCNVVVKPLCIKICAPGCVCRGLGYLRNKKKVCVPRSKC 66
DB 11 GRG-CGQNEEFTQGSACEPSC-NRPRAQAQCTLCQICVGCRCQGLFNSSGRGCVTPREC 67

```

Query Match      27.6%; Score 108; DB 5; Length 84;
Best Local Similarity 37.7%; Pred. No. 3 4e-06;
Matches 26; Conservative 7; Mismatches 30; Indels 6; Gaps 3;

QY 3 FGGLG--GRGKCPSEIFSRCDGRQCFNVP--KPLCIKICAPGCVRLGYLRNKK 57
DB 13 FAFGLVSANAKCGEIIYQRCGTACERTCSNGEWNKP-CKQPCVDKCFQGRGLRDN 71
QY 58 KVCVPRSKC 66
DB 72 GNCVRAWRC 80

RESULT 14
Q9VU94 PRELIMINARY; PRT; 3843 AA.
ID Q9VU94
AC Q9VU94;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hml protein.
GN HML OR G07002.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=107311132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Anantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Gallo R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR HSSP; P56682; AAF49795.2; -.
DR FlyBase; FBgn0029167; Hml.
DR InterPro; IPR002557; Chitin_bind_PeR.
Query Match      28.0%; Score 109.5; DB 5; Length 949;
Best Local Similarity 39.0%; Pred. No. 2e-05;
Matches 23; Conservative 7; Mismatches 24; Indels 5; Gaps 3;

QY 12 CPSEIFSRCDGRQCFNVPKPLC-IKICAPGCVRLGYLR--NKKKVCVPRSKC 66
DB 229 CGANEQYSACFSQCPSQCD-PSTPACFAPGCPGICLPFYIRDDSSPSACVPRGLC 286

RESULT 13
Q816Q5 PRELIMINARY; PRT; 84 AA.
ID Q816Q5
AC Q816Q5;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Putative salivary secreted serine protease inhibitor.
OS Anopheles stephensi (Indo-Pakistan malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
OX NCBI_TaxID=30069;
RN [1]
RP SEQUENCE FROM N.A.
RC TTSUE=Salivary gland;
RA Valenzuela J.G., Francischetti I.M.B., Pham V.M., Garfield M.K.,
RA Ribeiro J.M.C.;
RT "Exploring the sialome of the mosquito, Anopheles stephensi."
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY162234; AAO06826.1; -.
KW Protease.
SQ SEQUENCE 84 AA; 9369 MW; CF34128FDDDF64C CRC64;
```

```
DR InterPro; IPR006207; Cys knot C.
DR InterPro; IPR006209; EGF_like_C.
DR InterPro; IPR000421; FAS5_C.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR002172; LDL_receptor A.
DR InterPro; IPR000436; Sushi_SCR_CCF.
DR InterPro; IPR002919; TIL Cysrich.
DR InterPro; IPR001007; VWF_C.
DR InterPro; IPR001846; VWF_D.
DR Pfam; PF01607; CEM_14; 1.
DR Pfam; PF00754; F5_F8_type_C; 2.
DR Pfam; PF01826; TIL; 4.
DR Pfam; PF00094; vwd; 5.
DR SMART; SM00032; CCP; 1.
DR SMART; SM00494; ChtBD2; 1.
DR SMART; SM00041; CT; 1.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00231; FAS5C; 2.
DR SMART; SM00192; LDLa; 1.
DR SMART; SM00214; VWC; 3.
DR SMART; SM00216; VWD; 5.
DR PROSITE; PS01185; CTCK_1; 1.
DR PROSITE; PS01225; CTCK_2; 1.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01286; FAS5C_2; 1.
DR PROSITE; PS50068; LDLRA_2; 1.
DR PROSITE; PS01208; VWF; 1.
DR EGF-like domain.
KW
SQ SEQUENCE 3843 AA; 426241 MW; 78D770C0B8B6319D CRC64;

Query Match 27.6%; Score 108; DB 5; Length 3843;
Best Local Similarity 33.9%; Pred. No. 0.00011;
Matches 19; Conservative 7; Mismatches 30; Indels 0; Gaps 0;

Qy 11 KCPSEIFSRCDGRCQRFPCPNVVPKPLCIKICAPGCVCLGYLRNKKKVCVPRSKC 66
Db 1130 KCPLGQVFDECGDGCALSCDDLPSKSGCKRECVGCRCPHGVEYVNEDEGCVPKKMC 1185

RESULT 15
Q9U5D0
ID Q9U5D0 PRELIMINARY; PRT; 3843 AA.
AC Q9U5D0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DI 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
ET 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE HEMOLECTIN.
GN HML OR CG7002.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]_
RP SEQUENCE FROM N.A.
RA Goto A., Kumagai C., Kumagai T., Hirose J., Narita H., Beck K.,
RA Mori H., Kadowaki T., Kitagawa Y.;
RT "Identification and cloning of Hemoclectin : A Drosophila protein with
RT homology to the silkworm hemocytin and mammalian von Willebrand
RT factor.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB035891; BAA88518.1; -.
DR HSP; F56682; ICCV.
DR FlyBase; FBgn0029167; Hml.
DR InterPro; IPR002557; Chitin bind PerA.
DR InterPro; IPR006207; Cys knot C.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR000421; FAS5_C.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR002172; LDL_receptor A.
DR InterPro; IPR000436; Sushi_SCR_CCF.
DR InterPro; IPR002919; TIL_Cysrich.
```

```
DR InterPro; IPR001007; VWF_C.
DR InterPro; IPR001846; VWF_D.
DR Pfam; PF01607; CEM_14; 1.
DR Pfam; PF00754; F5_F8_type_C; 2.
DR Pfam; PF01826; TIL; 4.
DR Pfam; PF00094; vwd; 5.
DR SMART; SM00032; CCP; 1.
DR SMART; SM00494; ChtBD2; 1.
DR SMART; SM00041; CT; 1.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00231; FAS5C; 2.
DR SMART; SM00192; LDLa; 1.
DR SMART; SM00214; VWC; 3.
DR SMART; SM00216; VWD; 5.
DR PROSITE; PS01185; CTCK_1; 1.
DR PROSITE; PS01225; CTCK_2; 1.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01286; FAS5C_2; 1.
DR PROSITE; PS50068; LDLRA_2; 1.
DR PROSITE; PS01208; VWF; 1.
DR EGF-like domain.
KW
SQ SEQUENCE 3843 AA; 426349 MW; 20095BC3BF21E49D CRC64;

Query Match 27.6%; Score 108; DB 5; Length 3843;
Best Local Similarity 33.9%; Pred. No. 0.00011;
Matches 19; Conservative 7; Mismatches 30; Indels 0; Gaps 0;

Qy 11 KCPSEIFSRCDGRCQRFPCPNVVPKPLCIKICAPGCVCLGYLRNKKKVCVPRSKC 66
Db 1130 KCPLGQVFDECGDGCALSCDDLPSKSGCKRECVGCRCPHGVEYVNEDEGCVPKKMC 1185

Search completed: November 17, 2003, 08:09:18
Job time : 36 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2003, 08:08:37 ; Search time 20 Seconds
(without alignments)
162,520 Million cell updates/sec

Title: US-09-506-978-1

Perfect score: 391

Sequence: 1 GGFGLGGRGKCPSEIFSR.....CRGLYLNRKKVCVPRSKG 67

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 256763 seqs, 48513562 residues

Total number of hits satisfying chosen parameters: 256763

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents_AA_New:*

- 1: /cgn2_6/prodata/2/paa/PCT_NEW_COMB.pep.*
- 2: /cgn2_6/prodata/2/paa/US06_NEW_COMB.pep.*
- 3: /cgn2_6/prodata/2/paa/US07_NEW_COMB.pep.*
- 4: /cgn2_6/prodata/2/paa/US08_NEW_COMB.pep.*
- 5: /cgn2_6/prodata/2/paa/US09_NEW_COMB.pep.*
- 6: /cgn2_6/prodata/2/paa/US10_NEW_COMB.pep.*
- 7: /cgn2_6/prodata/2/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	80	20.5	1373	6	US-10-662-613-2
2	72	18.4	366	5	US-09-963-693A-103
3	72	18.4	366	5	US-09-205-658A-103
4	72	18.4	941	1	PCT-US03-28227-4736
5	72	18.4	949	1	PCT-US03-28227-4735
6	72	18.4	956	7	US-60-487-610-1489
7	72	18.4	958	1	PCT-US03-28227-4734
8	72	18.4	968	1	PCT-US03-28227-4733
9	72	18.4	1359	6	US-10-388-838-54
10	72	18.4	1360	6	US-10-388-838-50
11	72	18.4	1367	1	PCT-US03-11867-10
12	72	18.4	1367	6	US-10-365-352-98
13	72	18.4	1367	6	US-10-443-466A-19
14	72	18.4	1367	7	US-60-487-610-1518
15	72	18.4	1367	7	US-60-490-890-892
16	72	18.4	1367	7	US-60-500-337-1579
17	72	18.4	1367	7	US-60-505-218-345
18	72	18.4	1368	7	US-60-500-337-1580
19	72	18.4	1368	7	US-60-505-218-346
20	71	18.2	771	1	PCT-US03-26780-3242
21	71	18.2	1587	7	US-60-487-610-1941
22	70	17.9	1350	1	PCT-US03-33491-70
23	69	17.6	816	1	PCT-US03-26780-3243
24	68	17.4	47	5	US-09-714-602-400
25	67.5	17.3	661	6	US-10-473-040-707
26	67.5	17.3	751	6	US-10-679-063-22635

27	67.5	17.3	1269	1	PCT-US03-26780-3187	Sequence 3187, Ap
28	67	17.1	1037	1	PCT-US03-30894-36	Sequence 36, Appl
29	66.5	17.0	102	1	PCT-US03-30720-2447	Sequence 2447, Ap
30	66.5	17.0	105	1	PCT-US03-30720-1215	Sequence 1215, Ap
31	66.5	17.0	467	1	PCT-US02-24483-176	Sequence 176, App
32	66.5	17.0	757	6	US-10-474-794-177	Sequence 177, App
33	66.5	17.0	757	6	US-10-474-794-196	Sequence 196, App
34	66.5	17.0	870	1	PCT-US03-26780-2644	Sequence 2644, Ap
35	66.5	17.0	1435	1	PCT-US02-24483-54	Sequence 54, Appl
36	66.5	17.0	2140	6	US-10-679-063-23340	Sequence 23340, A
37	66	16.9	373	6	US-10-473-305-1339	Sequence 1339, Ap
38	66	16.9	556	1	PCT-US03-26232-9	Sequence 9, Appli
39	66	16.9	994	1	PCT-US03-28227-4983	Sequence 4983, Ap
40	66	16.9	1037	1	PCT-US03-28227-4984	Sequence 4984, Ap
41	65.5	16.8	400	7	US-60-500-337-2492	Sequence 2492, Ap
42	65.5	16.8	689	6	US-10-612-699-5	Sequence 5, Appli
43	65.5	16.8	903	6	US-10-357-885-104	Sequence 104, App
44	65.5	16.8	1138	5	US-09-623-063A-2	Sequence 2, Appli
45	65.5	16.8	1138	7	US-60-500-337-2416	Sequence 2416, Ap

ALIGNMENTS

RESULT 1

US-10-662-613-2

; Sequence 2, Application US/10662613

; GENERAL INFORMATION:

; APPLICANT: Farid, Hossain A.

; APPLICANT: Otieno, Charles J.

; APPLICANT: Benkel, Bernhard F.

; TITLE OF INVENTION: Insulin-Like Growth Factor-1 Receptor (IGF-1R) Polymorphic

; TITLE OF INVENTION: Alleles and Use of the Same to Identify DNA Markers for

; TITLE OF INVENTION: Reproductive Longevity

; FILE REFERENCE: P05562US00

; CURRENT APPLICATION NUMBER: US/10/662,613

; NUMBER OF SEQ ID NOS: 22

; SOFTWARE: Patent in version 3.2

; SEQ ID NO 2

; LENGTH: 1373

; TYPE: PRT

; ORGANISM: Mus musculus

US-10-662-613-2

Query Match	20.5%;	Score 80;	DB 6;	Length 1373;
Best Local Similarity	33.8%;	Pred. No. 44;		
Matches	23;	Conservative	5;	Mismatches 18; Indels 22; Gaps 4;
QY	14	SNEIFSRC--DGRCORFCNVPKPLCI---KICAPGCV-----	CRGLYLNR 54	
DB	198	NNEYNYRCWTTNRCQMCPSVCGKRACTENNECCHECLGSCHTPDNTTCVACRHHY- 256		
QY	55	NKKKVCVP 62		
DB	257	--KGVCVP 262		

RESULT 2

US-09-963-693A-103

; Sequence 103, Application US/09963693A

; GENERAL INFORMATION:

; APPLICANT: Ruvkun, Gary

; APPLICANT: Ogg, Scott

; TITLE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS

; FILE REFERENCE: 00786/351004

; CURRENT APPLICATION NUMBER: US/09/963,693A

; PRIOR FILING DATE: 2001-09-25

; PRIOR APPLICATION NUMBER: 09/205,658

; PRIOR FILING DATE: 1998-12-03

; PRIOR APPLICATION NUMBER: 08/857,076

; PRIOR FILING DATE: 1997-05-15


```

; PRIOR APPLICATION NUMBER: 08/888,534
; PRIOR FILING DATE: 1997-07-07
; PRIOR APPLICATION NUMBER: US98/10080
; PRIOR FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 328
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 103
; LENGTH: 366
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-963-693A-103

```

```

Query Match      18.4%; Score 72; DB 5; Length 366;
Best Local Similarity 30.9%; Pred. No. 60;
Matches 21; Conservative 5; Mismatches 20; Indels 22; Gaps 4;

QY 14 SNEIFSRC--DGRQRCFCNVVPKPLCI---KICAPGCV-----CRLGYLR 54
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
61 NNEYNRCWTTNRCQKMCPCSTCGKRACTENNECCHEPCLGSCSPDNDTACVACRHYYYA 120
QY 55 NKKKVCVP 62
Db |||||
121 G---VCVP 125

```

```

RESULT 3
US-09-205-658A-103
; Sequence 103, Application US/09205658A
; GENERAL INFORMATION:
; APPLICANT: Ruvkun, Gary
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
; FILE REFERENCE: 00786/351004
; CURRENT APPLICATION NUMBER: US/09/205,658A
; CURRENT FILING DATE: 1998-12-03
; PRIOR APPLICATION NUMBER: 08/857,076
; PRIOR FILING DATE: 1997-05-15
; PRIOR APPLICATION NUMBER: 08/888,534
; PRIOR FILING DATE: 1997-07-07
; PRIOR APPLICATION NUMBER: US98/10080
; PRIOR FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 331
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 103
; LENGTH: 366
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-205-658A-103

```

```

Query Match      18.4%; Score 72; DB 5; Length 366;
Best Local Similarity 30.9%; Pred. No. 60;
Matches 21; Conservative 5; Mismatches 20; Indels 22; Gaps 4;

QY 14 SNEIFSRC--DGRQRCFCNVVPKPLCI---KICAPGCV-----CRLGYLR 54
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
61 NNEYNRCWTTNRCQKMCPCSTCGKRACTENNECCHEPCLGSCSPDNDTACVACRHYYYA 120
QY 55 NKKKVCVP 62
Db |||||
121 G---VCVP 125

```

```

RESULT 4
PCT-US03-28227-4736
; Sequence 4736, Application PC/TUS0328227
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION; SCHMIDT, Jeanette P.;
; APPLICANT: WRIGHT, Rachel J.; BRUNS, Christopher M.;
; APPLICANT: MARIANOVIC, Mirjana M.; SHEN, Fan;
; APPLICANT: HARTSHORNE, Toinette A.; SUCHOROLSKI, Martin;
; APPLICANT: ALTUS, Christina M.; PITTS, Steven J.;
; APPLICANT: ELDER, Linda V.; MOONEY, Elizabeth M.;

```

```

; APPLICANT: DELEGANE, Angelo M.; PANESAR, Iqbal S.;
; APPLICANT: BANVILLE, Steven C.; REDDY, Thirupathi P.;
; APPLICANT: STEVENS, Kristian A.; BLANCHARD, John L.;
; APPLICANT: PANZER, Scott R.; WANG, Xinhao;
; APPLICANT: AU, Alan P.; GERSTIN, Edward H., Jr.;
; APPLICANT: PERALTA, Careyna H.; ANDERSON, Scott E.;
; APPLICANT: RIOUX, Pierre; SHEN, Edward J.;
; APPLICANT: WU, Mingham C.; STUVE, Laura L.;
; APPLICANT: LAGACE, Robert E.; SPIRO, Peter A.;
; APPLICANT: STEWART, Elizabeth A.; WINGROVE, James A.;
; APPLICANT: VITT, Ursula A.; KIRTON, Edward;
; APPLICANT: XU, Yuming; KWONG, Mary;
; APPLICANT: POLICKY, Jennifer L.; HURWITZ, Bonnie L.;
; APPLICANT: MA, Yan; JACKSON, Jennifer L.;
; APPLICANT: GIETZEN, Darryl; PATURY, Srikanth;
; APPLICANT: SHI, Xiaobing; SUAREZ, Charlyn J.
; TITLE OF INVENTION: MOLECULES FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PN-0100 PCT
; CURRENT APPLICATION NUMBER: PCT/US03/28227
; CURRENT FILING DATE: 2003-09-12
; PRIOR APPLICATION NUMBER: US 60/410,260
; PRIOR FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 60/410,259
; PRIOR FILING DATE: 2002-09-12
; NUMBER OF SEQ ID NOS: 5444
; SOFTWARE: PERL Program
; SEQ ID NO 4736
; LENGTH: 941
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; FEATURE:
; OTHER INFORMATION: Incyte ID No: 962818.PT196p
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1) ... (941)
; OTHER INFORMATION: unknown or other
PCT-US03-28227-4736

```

```

Query Match      18.4%; Score 72; DB 1; Length 941;
Best Local Similarity 33.3%; Pred. No. 1.3e+02;
Matches 20; Conservative 4; Mismatches 18; Indels 18; Gaps 4;

QY 7 GGRGKCPSEIFSRCDGRQRCFCNVVPKPLCIKICAPGCVCRLGYLENKKKVCVPSK 66
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
362 GNGGCDNSI-----CINTVGS----FK-CGP---CRLGFLGNQSGCLIPARTC 403

```

```

RESULT 5
PCT-US03-28227-4735
; Sequence 4735, Application PC/TUS0328227
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION; SCHMIDT, Jeanette P.;
; APPLICANT: WRIGHT, Rachel J.; BRUNS, Christopher M.;
; APPLICANT: MARIANOVIC, Mirjana M.; SHEN, Fan;
; APPLICANT: HARTSHORNE, Toinette A.; SUCHOROLSKI, Martin;
; APPLICANT: ALTUS, Christina M.; PITTS, Steven J.;
; APPLICANT: ELDER, Linda V.; MOONEY, Elizabeth M.;
; APPLICANT: DELEGANE, Angelo M.; PANESAR, Iqbal S.;
; APPLICANT: BANVILLE, Steven C.; REDDY, Thirupathi P.;
; APPLICANT: STEVENS, Kristian A.; BLANCHARD, John L.;
; APPLICANT: PANZER, Scott R.; WANG, Xinhao;
; APPLICANT: AU, Alan P.; GERSTIN, Edward H., Jr.;
; APPLICANT: PERALTA, Careyna H.; ANDERSON, Scott E.;
; APPLICANT: RIOUX, Pierre; SHEN, Edward J.;
; APPLICANT: WU, Mingham C.; STUVE, Laura L.;
; APPLICANT: LAGACE, Robert E.; SPIRO, Peter A.;
; APPLICANT: STEWART, Elizabeth A.; WINGROVE, James A.;
; APPLICANT: VITT, Ursula A.; KIRTON, Edward;
; APPLICANT: XU, Yuming; KWONG, Mary;
; APPLICANT: POLICKY, Jennifer L.; HURWITZ, Bonnie L.;
; APPLICANT: MA, Yan; JACKSON, Jennifer L.;
; APPLICANT: GIETZEN, Darryl; PATURY, Srikanth;

```

```

; APPLICANT: SHI, Xiaobing; SUAREZ, Charlyn J.
; TITLE OF INVENTION: MOLECULES FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PN-0100 PCT
; CURRENT APPLICATION NUMBER: PCT/US03/28227
; CURRENT FILING DATE: 2003-09-12
; PRIOR APPLICATION NUMBER: US 60/410,260
; PRIOR FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 60/410,259
; PRIOR FILING DATE: 2002-09-12
; NUMBER OF SEQ ID NOS: 5444
; SOFTWARE: PERL Program
; SEQ ID NO 4735
; LENGTH: 949
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 962818.PT195p
; LOCATION: (1) ... (949)
; OTHER INFORMATION: unknown or other
PCT-US03-28227-4735

Query Match 18.4%; Score 72; DB 1; Length 949;
Best Local Similarity 33.3%; Pred. No. 1.3e+02;
Matches 20; Conservative 4; Mismatches 18; Indels 18; Gaps 4;

QY 7 GGGKCPSEIFSRCDGRCQRCFNVVPKPLCIKICAPGCVCLGYLRNKKVCVPRSKC 66
DB 370 GNNGGCDPNSI-----CTNTVGS---FK-CGP---CRLGFLGNQSQGCLPARTC 411

RESULT 6
US-60-487-610-1489
; Sequence 1489, Application US/60487610
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: HUANG, Honglin
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: LIVER FIBROSIS IN HEPATITIS C VIRUS-INFECTED SUBJECTS,
; TITLE OF INVENTION: METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001469
; CURRENT APPLICATION NUMBER: US/60/487,610
; CURRENT FILING DATE: 2003-07-17
; NUMBER OF SEQ ID NOS: 97101
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1489
; LENGTH: 956
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-487-610-1489

Query Match 18.4%; Score 72; DB 7; Length 956;
Best Local Similarity 33.3%; Pred. No. 1.3e+02;
Matches 20; Conservative 4; Mismatches 18; Indels 18; Gaps 4;

QY 7 GGGKCPSEIFSRCDGRCQRCFNVVPKPLCIKICAPGCVCLGYLRNKKVCVPRSKC 66
DB 377 GNNGGCDPNSI-----CTNTVGS---FK-CGP---CRLGFLGNQSQGCLPARTC 418

RESULT 7
PCT-US03-28227-4734
; Sequence 4734, Application PC/TUS0328227
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION; SCHMIDT, Jeanette P.;
; APPLICANT: WRIGHT, Rachel J.; BRUNS, Christopher M.;
; APPLICANT: MARJANOVIC, Mirjana M.; SHEN, Fan;
; APPLICANT: HARTSHORNE, Toinette A.; SUCHOROLSKI, Martin;
; APPLICANT: ALTUS, Christina M.; PITTS, Steven J.;
; APPLICANT: ELDER, Linda V.; MOONEY, Elizabeth M.;
; APPLICANT: DELEGEANE, Angelo M.; PANESAR, Iqbal S.;
; APPLICANT: BANVILLE, Steven C.; REDDY, Thirupathi P.;
; APPLICANT: STEVENS, Kristian A.; BLANCHARD, John L.;
; APPLICANT: PANZER, Scott R.; WANG, Xinhao;
; APPLICANT: AU, Alan P.; GERSTIN, Edward H., Jr.;
; APPLICANT: PERALTA, Careyna H.; ANDERSON, Scott E.;
; APPLICANT: RIOUX, Pierre; SHEN, Edward J.;
; APPLICANT: WU, Mingham C.; STUVE, Laura L.;
; APPLICANT: LAGACE, Robert E.; SPIRO, Peter A.;
; APPLICANT: STEWART, Elizabeth A.; WINGROVE, James A.;
; APPLICANT: VITT, Ursula A.; KIRTON, Edward;
; APPLICANT: XU, Yuming; KWONG, Mary;
; APPLICANT: POLICKY, Jennifer L.; HURWITZ, Bonnie L.;
; APPLICANT: MA, Yan; JACKSON, Jennifer L.;
; APPLICANT: GIETZEN, Darryl; PATURY, Srikanth;
; APPLICANT: SHI, Xiaobing; SUAREZ, Charlyn J.
; TITLE OF INVENTION: MOLECULES FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PN-0100 PCT
; CURRENT APPLICATION NUMBER: PCT/US03/28227
; CURRENT FILING DATE: 2003-09-12

```

```

; APPLICANT: BANVILLE, Steven C.; REDDY, Thirupathi P.;
; APPLICANT: STEVENS, Kristian A.; BLANCHARD, John L.;
; APPLICANT: PANZER, Scott R.; WANG, Xinhao;
; APPLICANT: AU, Alan P.; GERSTIN, Edward H., Jr.;
; APPLICANT: PERALTA, Careyna H.; ANDERSON, Scott E.;
; APPLICANT: RIOUX, Pierre; SHEN, Edward J.;
; APPLICANT: WU, Mingham C.; STUVE, Laura L.;
; APPLICANT: LAGACE, Robert E.; SPIRO, Peter A.;
; APPLICANT: STEWART, Elizabeth A.; WINGROVE, James A.;
; APPLICANT: VITT, Ursula A.; KIRTON, Edward;
; APPLICANT: XU, Yuming; KWONG, Mary;
; APPLICANT: POLICKY, Jennifer L.; HURWITZ, Bonnie L.;
; APPLICANT: MA, Yan; JACKSON, Jennifer L.;
; APPLICANT: GIETZEN, Darryl; PATURY, Srikanth;
; APPLICANT: SHI, Xiaobing; SUAREZ, Charlyn J.
; TITLE OF INVENTION: MOLECULES FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PN-0100 PCT
; CURRENT APPLICATION NUMBER: PCT/US03/28227
; CURRENT FILING DATE: 2003-09-12
; PRIOR APPLICATION NUMBER: US 60/410,260
; PRIOR FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 60/410,259
; PRIOR FILING DATE: 2002-09-12
; NUMBER OF SEQ ID NOS: 5444
; SOFTWARE: PERL Program
; SEQ ID NO 4734
; LENGTH: 958
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 962818.PT193p
PCT-US03-28227-4734

Query Match 18.4%; Score 72; DB 1; Length 958;
Best Local Similarity 33.3%; Pred. No. 1.4e+02;
Matches 20; Conservative 4; Mismatches 18; Indels 18; Gaps 4;

QY 7 GGGKCPSEIFSRCDGRCQRCFNVVPKPLCIKICAPGCVCLGYLRNKKVCVPRSKC 66
DB 379 GNNGGCDPNSI-----CTNTVGS---FK-CGP---CRLGFLGNQSQGCLPARTC 420

RESULT 8
PCT-US03-28227-4733
; Sequence 4733, Application PC/TUS0328227
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION; SCHMIDT, Jeanette P.;
; APPLICANT: WRIGHT, Rachel J.; BRUNS, Christopher M.;
; APPLICANT: MARJANOVIC, Mirjana M.; SHEN, Fan;
; APPLICANT: HARTSHORNE, Toinette A.; SUCHOROLSKI, Martin;
; APPLICANT: ALTUS, Christina M.; PITTS, Steven J.;
; APPLICANT: ELDER, Linda V.; MOONEY, Elizabeth M.;
; APPLICANT: DELEGEANE, Angelo M.; PANESAR, Iqbal S.;
; APPLICANT: BANVILLE, Steven C.; REDDY, Thirupathi P.;
; APPLICANT: STEVENS, Kristian A.; BLANCHARD, John L.;
; APPLICANT: PANZER, Scott R.; WANG, Xinhao;
; APPLICANT: AU, Alan P.; GERSTIN, Edward H., Jr.;
; APPLICANT: PERALTA, Careyna H.; ANDERSON, Scott E.;
; APPLICANT: RIOUX, Pierre; SHEN, Edward J.;
; APPLICANT: WU, Mingham C.; STUVE, Laura L.;
; APPLICANT: LAGACE, Robert E.; SPIRO, Peter A.;
; APPLICANT: STEWART, Elizabeth A.; WINGROVE, James A.;
; APPLICANT: VITT, Ursula A.; KIRTON, Edward;
; APPLICANT: XU, Yuming; KWONG, Mary;
; APPLICANT: POLICKY, Jennifer L.; HURWITZ, Bonnie L.;
; APPLICANT: MA, Yan; JACKSON, Jennifer L.;
; APPLICANT: GIETZEN, Darryl; PATURY, Srikanth;
; APPLICANT: SHI, Xiaobing; SUAREZ, Charlyn J.
; TITLE OF INVENTION: MOLECULES FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PN-0100 PCT
; CURRENT APPLICATION NUMBER: PCT/US03/28227
; CURRENT FILING DATE: 2003-09-12

```

```
/ PRIOR APPLICATION NUMBER: US 60/410,260
/ PRIOR FILING DATE: 2002-09-12
/ PRIOR APPLICATION NUMBER: US 60/410,259
/ PRIOR FILING DATE: 2002-09-12
/ NUMBER OF SEQ ID NOS: 5444
/ SOFTWARE: PERL Program
/ SEQ ID NO 4733
/ LENGTH: 968
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc feature
/ OTHER INFORMATION: Incyte ID No: 962818.PT191p
/ NAME/KEY: unsure
/ LOCATION: (1)... (968)
/ OTHER INFORMATION: unknown or other
PCT-US03-28227-4733

Query Match      18.4%; Score 72; DB 1; Length 968;
Best Local Similarity 33.3%; Pred. No. 1.4e+02;
Matches 20; Conservative 4; Mismatches 18; Indels 18; Gaps 4;

Qy 7 GGRGKCSNEIFSRCDGRQCFNVPKPLCIKICAPGCVRLGYLNNKKKVCVPSK 66
Db 389 GNNGGCDPNSI-----CTNVGS---FK-CGP---CRLGFLGNQSGCLPARTC 430

RESULT 9
US-10-388-838-54
/ Sequence 54, Application US/10388838
/ GENERAL INFORMATION:
/ APPLICANT: David W. Morris
/ APPLICANT: Marc Malandro
/ TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
/ FILE REFERENCE: 529452001600
/ CURRENT APPLICATION NUMBER: US/10/388,838
/ CURRENT FILING DATE: 2003-03-14
/ NUMBER OF SEQ ID NOS: 114
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 54
/ LENGTH: 1359
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-388-838-54

Query Match      18.4%; Score 72; DB 6; Length 1359;
Best Local Similarity 30.9%; Pred. No. 1.8e+02;
Matches 21; Conservative 5; Mismatches 20; Indels 22; Gaps 4;

Qy 14 SNEIFSRC--DGRQCFNVPKPLCI---KICAPGCV-----CRLGYLR 54
Db 196 NNEYNRCWTNRCQKMPSTCGKRACTENNECHPECLGSCSAPDNDTACVACRHYIA 255

Qy 55 NKKKVCVP 62
Db 256 G---VCVP 260

RESULT 10
US-10-388-838-50
/ Sequence 50, Application US/10388838
/ GENERAL INFORMATION:
/ APPLICANT: David W. Morris
/ APPLICANT: Marc Malandro
/ TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
/ FILE REFERENCE: 529452001600
/ CURRENT APPLICATION NUMBER: US/10/388,838
/ CURRENT FILING DATE: 2003-03-14
/ NUMBER OF SEQ ID NOS: 114
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 50
/ LENGTH: 1360
```

```
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-388-838-50

Query Match      18.4%; Score 72; DB 6; Length 1360;
Best Local Similarity 30.9%; Pred. No. 1.8e+02;
Matches 21; Conservative 5; Mismatches 20; Indels 22; Gaps 4;

Qy 14 SNEIFSRC--DGRQCFNVPKPLCI---KICAPGCV-----CRLGYLR 54
Db 196 NNEYNRCWTNRCQKMPSTCGKRACTENNECHPECLGSCSAPDNDTACVACRHYIA 255

Qy 55 NKKKVCVP 62
Db 256 G---VCVP 260

RESULT 11
PCT-US03-11867-10
/ Sequence 10, Application PC/TUS0311867
/ GENERAL INFORMATION:
/ APPLICANT: Hitoshi, Yasumichi
/ APPLICANT: Jenkins, Yonchu
/ APPLICANT: Rigel Pharmaceuticals, Inc.
/ TITLE OF INVENTION: Methods of Assaying for Cell Cycle Modulators
/ FILE REFERENCE: 021044-003310PC
/ CURRENT APPLICATION NUMBER: PCT/US03/11867
/ CURRENT FILING DATE: 2003-04-15
/ PRIOR APPLICATION NUMBER: US 10/123,568
/ PRIOR FILING DATE: 2002-04-15
/ PRIOR APPLICATION NUMBER: US 10/123,731
/ PRIOR FILING DATE: 2002-04-15
/ PRIOR APPLICATION NUMBER: US 60/373,366
/ PRIOR FILING DATE: 2002-04-16
/ NUMBER OF SEQ ID NOS: 60
/ SOFTWARE: Patent In Ver. 2.1
/ SEQ ID NO 10
/ LENGTH: 1367
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ OTHER INFORMATION: Insulin-like growth factor 1 receptor (IGF1R)
/ OTHER INFORMATION: (cell surface kinase receptor)
PCT-US03-11867-10

Query Match      18.4%; Score 72; DB 1; Length 1367;
Best Local Similarity 30.9%; Pred. No. 1.8e+02;
Matches 21; Conservative 5; Mismatches 20; Indels 22; Gaps 4;

Qy 14 SNEIFSRC--DGRQCFNVPKPLCI---KICAPGCV-----CRLGYLR 54
Db 198 NNEYNRCWTNRCQKMPSTCGKRACTENNECHPECLGSCSAPDNDTACVACRHYIA 257

Qy 55 NKKKVCVP 62
Db 258 G---VCVP 262

RESULT 12
US-10-365-352-98
/ Sequence 98, Application US/10365352
/ GENERAL INFORMATION:
/ APPLICANT: Wraight, Christopher
/ APPLICANT: Werther, Georger
/ APPLICANT: Dean, Nicholas
/ APPLICANT: Dobie, Kenneth
/ TITLE OF INVENTION: Modulation of insulin-like growth factor I receptor expressic
/ FILE REFERENCE: 229752002500
/ CURRENT APPLICATION NUMBER: US/10/365,352
/ CURRENT FILING DATE: 2003-02-11
/ NUMBER OF SEQ ID NOS: 98
/ SOFTWARE: Patent In version 3.1
/ SEQ ID NO 98
```

LENGTH: 1367
TYPE: PRT
ORGANISM: human
US-10-365-352-98

Query Match 18.4%; Score 72; DB 6; Length 1367;
Best Local Similarity 30.9%; Pred. No. 1.8e+02;
Matches 21; Conservative 5; Mismatches 20; Indels 22; Gaps 4;

QY 14 SNEIFSRC--DGRQRCFCPNVVPKPLCI---KICAPGCV-----CRLGYLR 54
Db 198 NNEYNVRCWTTNRCQKMCPSCTGKRACTENNECHPECLGSCSAPDNDTACVACRHYA 257
QY 55 NKKKVCVP 62
Db 258 G---VCVP 262

RESULT 13

US-10-443-466A-19
Sequence 19, Application US/10443466A
GENERAL INFORMATION:

APPLICANT: Wang, Yan
APPLICANT: Pachter, Jonathan A
APPLICANT: Hailey, Judith
APPLICANT: Greenberg, Robert
APPLICANT: Leonard, Presta
APPLICANT: Brans, Peter
APPLICANT: Feingersh, Diane
APPLICANT: Williams, Denise
APPLICANT: Srinivasan, Mohan
TITLE OF INVENTION: NEUTRALIZING HUMAN ANTI-IGFR ANTIBODY
FILE REFERENCE: OC01533-K-US

CURRENT APPLICATION NUMBER: US/10/443,466A
CURRENT FILING DATE: 2003-05-22
PRIOR APPLICATION NUMBER: 60/383,459
PRIOR FILING DATE: 2002-05-24
PRIOR APPLICATION NUMBER: 60/393,214
PRIOR FILING DATE: 2002-07-02
PRIOR APPLICATION NUMBER: 60/436,254
PRIOR FILING DATE: 2002-12-23
NUMBER OF SEQ ID NOS: 120
SOFTWARE: PatentIn version 3.1

SEQ ID NO 19
LENGTH: 1367
TYPE: PRT
ORGANISM: Homo sapiens
US-10-443-466A-19

Query Match 18.4%; Score 72; DB 6; Length 1367;
Best Local Similarity 30.9%; Pred. No. 1.8e+02;
Matches 21; Conservative 5; Mismatches 20; Indels 22; Gaps 4;

QY 14 SNEIFSRC--DGRQRCFCPNVVPKPLCI---KICAPGCV-----CRLGYLR 54
Db 198 NNEYNVRCWTTNRCQKMCPSCTGKRACTENNECHPECLGSCSAPDNDTACVACRHYA 257
QY 55 NKKKVCVP 62
Db 258 G---VCVP 262

RESULT 14

US-60-487-610-1518
Sequence 1518, Application US/60487610
GENERAL INFORMATION:

APPLICANT: CARGILL, Michele
APPLICANT: HUANG, Hongjin
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
LIVER FIBROSIS IN HEPATITIS C VIRUS-INFECTED SUBJECTS,
METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001469
CURRENT APPLICATION NUMBER: US/60/487,610

CURRENT FILING DATE: 2003-07-17
NUMBER OF SEQ ID NOS: 97101
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1518
LENGTH: 1367
TYPE: PRT
ORGANISM: Homo sapiens
US-60-487-610-1518

Query Match 18.4%; Score 72; DB 7; Length 1367;
Best Local Similarity 30.9%; Pred. No. 1.8e+02;
Matches 21; Conservative 5; Mismatches 20; Indels 22; Gaps 4;

QY 14 SNEIFSRC--DGRQRCFCPNVVPKPLCI---KICAPGCV-----CRLGYLR 54
Db 198 NNEYNVRCWTTNRCQKMCPSCTGKRACTENNECHPECLGSCSAPDNDTACVACRHYA 257
QY 55 NKKKVCVP 62
Db 258 G---VCVP 262

RESULT 15

US-60-490-890-892
Sequence 892, Application US/60490890
GENERAL INFORMATION:

APPLICANT: Li, Martha
APPLICANT: Rupnow, Brent A.
APPLICANT: Webster, Kevin R.
APPLICANT: Jackson, Donald
APPLICANT: Wong, Tai W.
TITLE OF INVENTION: BIOMARKERS OF CYCLIN-DEPENDENT KINASE MODULATION
FILE REFERENCE: D0310 PSP
CURRENT APPLICATION NUMBER: US/60/490,890
CURRENT FILING DATE: 2003-07-29
NUMBER OF SEQ ID NOS: 2779
SOFTWARE: PatentIn version 3.2
SEQ ID NO 892
LENGTH: 1367
TYPE: PRT
ORGANISM: Homo sapiens
US-60-490-890-892

Query Match 18.4%; Score 72; DB 7; Length 1367;
Best Local Similarity 30.9%; Pred. No. 1.8e+02;
Matches 21; Conservative 5; Mismatches 20; Indels 22; Gaps 4;

QY 14 SNEIFSRC--DGRQRCFCPNVVPKPLCI---KICAPGCV-----CRLGYLR 54
Db 198 NNEYNVRCWTTNRCQKMCPSCTGKRACTENNECHPECLGSCSAPDNDTACVACRHYA 257
QY 55 NKKKVCVP 62
Db 258 G---VCVP 262

Search completed: November 17, 2003, 08:14:05
Job time : 20 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2003, 08:07:17 ; Search time 22 Seconds
(without alignments)
128.856 Million cell updates/sec

Title: US-09-506-978-1
Perfect score: 391
Sequence: 1 GFGGLGGKGCPSNEIFSR.....CRLGLRNKKVCVPRSKCG 67

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5A COMB.pap.*
2: /cgn2_6/ptodata/1/iaa/5B COMB.pap.*
3: /cgn2_6/ptodata/1/iaa/6A COMB.pap.*
4: /cgn2_6/ptodata/1/iaa/6B COMB.pap.*
5: /cgn2_6/ptodata/1/iaa/PTCUS COMB.pap.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	391	100.0	92	4	US-09-394-630-2
2	107.5	27.5	2476	2	Sequence 2, Appli
3	106.5	27.2	2594	3	Sequence 2, Appli
4	106.5	27.2	5405	3	Sequence 7, Appli
5	100	25.6	84	2	Sequence 9, Appli
6	100	25.6	84	2	Sequence 59, Appli
7	100	25.6	84	2	Sequence 59, Appli
8	100	25.6	84	2	Sequence 59, Appli
9	100	25.6	84	2	Sequence 59, Appli
10	100	25.6	84	3	Sequence 59, Appli
11	100	25.6	84	3	Sequence 59, Appli
12	100	25.6	84	3	Sequence 59, Appli
13	100	25.6	84	3	Sequence 59, Appli
14	100	25.6	84	3	Sequence 59, Appli
15	100	25.6	84	3	Sequence 59, Appli
16	100	25.6	84	4	Sequence 59, Appli
17	100	25.6	91	2	Sequence 58, Appli
18	100	25.6	91	2	Sequence 128, App
19	100	25.6	91	2	Sequence 50, Appli
20	100	25.6	91	2	Sequence 128, App
21	100	25.6	91	2	Sequence 128, App
22	100	25.6	91	2	Sequence 128, App
23	100	25.6	91	2	Sequence 50, Appli
24	100	25.6	91	3	Sequence 128, App
25	100	25.6	91	3	Sequence 128, App
26	100	25.6	91	3	Sequence 128, App
27	100	25.6	91	3	Sequence 128, App

28 100 25.6 91 3 US-09-249-461-128 Sequence 128, App
29 100 25.6 91 3 US-09-249-448-128 Sequence 128, App
30 100 25.6 91 4 US-09-249-473-128 Sequence 128, App
31 98 25.1 115 4 US-09-219-383A-4 Sequence 1, Appli
32 98 25.1 2813 4 US-09-381-261A-1 Sequence 25, Appli
33 91 23.3 82 2 US-08-465-380-25 Sequence 49, Appli
34 91 23.3 82 2 US-08-465-380-49 Sequence 54, Appli
35 91 23.3 82 2 US-08-480-478-54 Sequence 25, Appli
36 91 23.3 82 2 US-08-486-397-25 Sequence 25, Appli
37 91 23.3 82 2 US-08-486-397-49 Sequence 25, Appli
38 91 23.3 82 2 US-08-486-399-25 Sequence 25, Appli
39 91 23.3 82 2 US-08-486-399-49 Sequence 25, Appli
40 91 23.3 82 2 US-08-461-965-25 Sequence 25, Appli
41 91 23.3 82 2 US-08-461-965-49 Sequence 49, Appli
42 91 23.3 82 2 US-08-326-110A-54 Sequence 54, Appli
43 91 23.3 82 2 US-08-634-641-25 Sequence 25, Appli
44 91 23.3 82 2 US-08-634-641-49 Sequence 49, Appli
45 91 23.3 82 3 US-09-249-471-25 Sequence 25, Appli

ALIGNMENTS

RESULT 1
US-09-394-630-2
; Sequence 2, Application US/09394630
; Patent No. 6395306
; GENERAL INFORMATION:
; APPLICANT: Cui, Xiangmin
; APPLICANT: Lu, Yuefeng
; TITLE OF INVENTION: Useful Properties of a Bee Venom Protein and Gene
; TITLE OF INVENTION: Encoding Same
; FILE REFERENCE: 019049-000200US
; CURRENT APPLICATION NUMBER: US/09/394.630
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: US 60/100,172
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 92
; TYPE: PRT
; ORGANISM: Apis mellifera
US-09-394-630-2

Query Match 100.0%; Score 391; DB 4; Length 92;
Best Local Similarity 100.0%; Pred. No. 1.5e-33;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GFGGLGGKGCPSNEIFSRCDGRCQRCFQNVVVKPLCIKICAPGCVCRGLYLNKKKVC 60
Db 26 GFGGLGGKGCPSNEIFSRCDGRCQRCFQNVVVKPLCIKICAPGCVCRGLYLNKKKVC 85
Qy 61 VPRSKCG 67
Db 86 VPRSKCG 92

RESULT 2
US-08-276-967-2
; Sequence 2, Application US/08276967
; Patent No. 5851817
; GENERAL INFORMATION:
; APPLICANT: Hardy, Daniel M.
; APPLICANT: Garbers, David L.
; TITLE OF INVENTION: Species-Specific Egg-Binding Proteins of
; TITLE OF INVENTION: Sperm
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P. O. Box 4433
; CITY: Houston

```
/ STATE: Texas
/ COUNTRY: USA
/ ZIP: 77210-4433
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/276,967
/ FILING DATE: Submitted Herewith
/ CLASSIFICATION: 530
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Kitchell, Barbara S.
/ REGISTRATION NUMBER: 33,928
/ REFERENCE/DOCKET NUMBER: UTSD:418\KIT
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 713-787-1400
/ TELEFAX: 713-789-2679
/ TELEX: 79-0924
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 2476 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-08-276-967-2

Query Match 27.5%; Score 107.5; DB 2; Length 2476;
Best Local Similarity 33.3%; Pred. No. 0.0048;
Matches 22; Conservative 9; Mismatches 18; Indels 17; Gaps 3;

QY 11 KCPSEIFSRCDGRC-----QRFCNVVPKLCIKICAPGCVCLGYLNKKKVCV 61
DB 1455 KCPGSSYSTCAMPATCLSLNPNPSCPSTLP-----CAEGCQCGHILSGTS-CV 1506

QY 62 PRSKCG 67
DB 1507 PLSQCG 1512

RESULT 3
US-08-718-388-7
; Sequence 7, Application US/08718388
; Patent No. 6271362
; GENERAL INFORMATION:
; APPLICANT: MORIKAWA, MINORU
; APPLICANT: HARADA, NAOKI
; TITLE OF INVENTION: GENE ENCODING IGG FC REGION-BINDING
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/718,388
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: MURPHY JR, GERALD M
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 0230-111
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5405 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
/ US-08-718-388-9

Query Match 27.2%; Score 106.5; DB 3; Length 5405;
Best Local Similarity 36.8%; Pred. No. 0.013;
Matches 21; Conservative 6; Mismatches 29; Indels 1; Gaps 1;

QY 11 KCPSEIFSRCDGRCQRCFPCNVVPKLCIKICAPGCVCLGYLNKKKVCVPRSKCG 67
DB 1531 ECPNPNHYELCADTCSLGCALSAPPQCDGCAEGCQCDSGFLYN-GQACVPIQCG 1586

RESULT 5
US-08-465-380-59
; Sequence 59, Application US/08465380
; Patent No. 5863894
```

```
/ TELEPHONE: (703) 205-8000
/ TELEFAX: (703) 205-8050
/ INFORMATION FOR SEQ ID NO: 7:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 2594 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-08-718-388-7

Query Match 27.2%; Score 106.5; DB 3; Length 2594;
Best Local Similarity 36.8%; Pred. No. 0.0064;
Matches 21; Conservative 6; Mismatches 29; Indels 1; Gaps 1;

QY 11 KCPSEIFSRCDGRCQRCFPCNVVPKLCIKICAPGCVCLGYLNKKKVCVPRSKCG 67
DB 1531 ECPNPNHYELCADTCSLGCALSAPPQCDGCAEGCQCDSGFLYN-GQACVPIQCG 1586

RESULT 4
US-08-718-388-9
; Sequence 9, Application US/08718388
; Patent No. 6271362
; GENERAL INFORMATION:
; APPLICANT: MORIKAWA, MINORU
; APPLICANT: HARADA, NAOKI
; TITLE OF INVENTION: GENE ENCODING IGG FC REGION-BINDING
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/718,388
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: MURPHY JR, GERALD M
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 0230-111
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5405 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
/ US-08-718-388-9

Query Match 27.2%; Score 106.5; DB 3; Length 5405;
Best Local Similarity 36.8%; Pred. No. 0.013;
Matches 21; Conservative 6; Mismatches 29; Indels 1; Gaps 1;

QY 11 KCPSEIFSRCDGRCQRCFPCNVVPKLCIKICAPGCVCLGYLNKKKVCVPRSKCG 67
DB 1531 ECPNPNHYELCADTCSLGCALSAPPQCDGCAEGCQCDSGFLYN-GQACVPIQCG 1586

RESULT 5
US-08-465-380-59
; Sequence 59, Application US/08465380
; Patent No. 5863894
```

GENERAL INFORMATION:
APPLICANT: George P. Vlasuk, Patric H. Stanssens,
APPLICANT: Joris H.L. Mensens, Marc J. Lauwereys,
APPLICANT: Yves R. Laroche, Laurent S. Jespers,
APPLICANT: Yannick G.J. Gansemans, Matthew Moyle,
APPLICANT: Peter W. Bergum
TITLE OF INVENTION: NEMATODE-EXTRACTED ANTICOAGULANT
TITLE OF INVENTION: PROTEIN
NUMBER OF SEQUENCES: 356
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/465,380
FILING DATE: June 5, 1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/326,110
FILING DATE: October 18, 1994
ATTORNEY/AGENT INFORMATION:
NAME: BIGGS, SUZANNE L.
REGISTRATION NUMBER: 30,158
REFERENCE/DOCKET NUMBER: 213/268
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 84 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Ancylostoma caninum
US-08-465-380-59

Query Match 25.6%; Score 100; DB 2; Length 84;
Best Local Similarity 30.4%; Pred. No. 0.0012;
Matches 21; Conservative 8; Mismatches 24; Indels 16; Gaps 3;

QY 11 KCPSEIFSRCDGR-CORFC-----PNVVKPLCIKICAPGCVCRGLYLNKK 57
Db 5 QCGENEKDYSCGSKCEDKKCKYDGVBEDEENPV---PCLVRVCHQDCVCEGFGYRNKD 61
QY 58 KVCVPRSKC 66
Db 62 DKCVSAEDC 70

RESULT 6
US-08-486-397-59
Sequence 59, Application US/08486397
Patent No. 5866542
GENERAL INFORMATION:
APPLICANT: George P. Vlasuk, Patric H. Stanssens,
APPLICANT: Joris H.L. Mensens, Marc J. Lauwereys,
APPLICANT: Yves R. Laroche, Laurent S. Jespers,
APPLICANT: Yannick G.J. Gansemans, Matthew Moyle,
APPLICANT: Peter W. Bergum
TITLE OF INVENTION: NEMATODE-EXTRACTED ANTICOAGULANT
TITLE OF INVENTION: PROTEIN

NUMBER OF SEQUENCES: 357
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/486,397
FILING DATE: June 5, 1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/326,110
FILING DATE: October 18, 1994
ATTORNEY/AGENT INFORMATION:
NAME: BIGGS, SUZANNE L.
REGISTRATION NUMBER: 30,158
REFERENCE/DOCKET NUMBER: 213/269
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 84 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Ancylostoma caninum
US-08-486-397-59

Query Match 25.6%; Score 100; DB 2; Length 84;
Best Local Similarity 30.4%; Pred. No. 0.0012;
Matches 21; Conservative 8; Mismatches 24; Indels 16; Gaps 3;

QY 11 KCPSEIFSRCDGR-CORFC-----PNVVKPLCIKICAPGCVCRGLYLNKK 57
Db 5 QCGENEKDYSCGSKCEDKKCKYDGVBEDEENPV---PCLVRVCHQDCVCEGFGYRNKD 61
QY 58 KVCVPRSKC 66
Db 62 DKCVSAEDC 70

RESULT 7
US-08-486-399-59
Sequence 59, Application US/08486399
Patent No. 5866543
GENERAL INFORMATION:
APPLICANT: George P. Vlasuk, Patric H. Stanssens,
APPLICANT: Joris H.L. Mensens, Marc J. Lauwereys,
APPLICANT: Yves R. Laroche, Laurent S. Jespers,
APPLICANT: Yannick G.J. Gansemans, Matthew Moyle,
APPLICANT: Peter W. Bergum
TITLE OF INVENTION: NEMATODE-EXTRACTED ANTICOAGULANT
TITLE OF INVENTION: PROTEIN
NUMBER OF SEQUENCES: 356
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.

```
/
/
/ ZIP: 90071
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
/ MEDIUM TYPE: storage
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: IBM P.C. DOS 5.0
/ SOFTWARE: Word Perfect 5.1
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/486,399
/ FILING DATE: June 5, 1995
/ CLASSIFICATION: 530
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/326,110
/ FILING DATE: October 18, 1994
/ ATTORNEY/AGENT INFORMATION:
/ NAME: BIGGS, SUZANNE L.
/ REGISTRATION NUMBER: 30,158
/ REFERENCE/DOCKET NUMBER: 213/270
/ TELEPHONE: (213) 489-1600
/ TELEFAX: (213) 955-0440
/ TELEX: 67-3510
/ INFORMATION FOR SEQ ID NO: 59:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 84 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: peptide
/ ORIGINAL SOURCE:
/ ORGANISM: Ancylostoma caninum
/
/ US-08-486-399-59
/
/
/ Query Match 25.6%; Score 100; DB 2; Length 84;
/ Best Local Similarity 30.4%; Pred. No. 0.0012;
/ Matches 21; Conservative 8; Mismatches 24; Indels 16; Gaps 3;
/
/ QY 11 KCPSEIFRCDGR-CQRF-----PNVVKPLCIKICAPGCVCRIGYLRNKK 57
/ Db 5 QCGENEKYDSGSKCKKCYDGVVEEDDEPNV---PCLVRVCHQDCVCEGFRNKD 61
/
/ QY 58 KVCVPRSKC 66
/ Db 62 DKCVSAEDC 70
/
/ RESULT 8
/ US-08-461-965-59
/ Sequence 59, Application US/08461965
/ Patent No. 5872098
/ GENERAL INFORMATION:
/ APPLICANT: George P. Vlasuk, Patric H. Stanssens,
/ APPLICANT: Joris H.L. Mensens, Marc J. Lauwereys,
/ APPLICANT: Yves R. Laroche, Laurent S. Jespers,
/ APPLICANT: Yannick G.J. Gansmans, Matthew Moyle,
/ APPLICANT: Peter W. Bergum
/ TITLE OF INVENTION: NEMATODE-EXTRACTED ANTICOAGULANT
/ NUMBER OF SEQUENCES: 356
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Lyon & Lyon
/ STREET: 633 West Fifth Street
/ CITY: Suite 4700
/ STATE: Los Angeles
/ COUNTRY: California
/ COUNTRY: U.S.A.
/ ZIP: 90071
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
/ MEDIUM TYPE: storage
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: IBM P.C. DOS 5.0
/ SOFTWARE: Word Perfect 5.1
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/486,399
/ FILING DATE: June 5, 1995
/ CLASSIFICATION: 530
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/326,110
/ FILING DATE: October 18, 1994
/ ATTORNEY/AGENT INFORMATION:
/ NAME: BIGGS, SUZANNE L.
/ REGISTRATION NUMBER: 30,158
/ REFERENCE/DOCKET NUMBER: 213/270
/ TELEPHONE: (213) 489-1600
/ TELEFAX: (213) 955-0440
/ TELEX: 67-3510
/ INFORMATION FOR SEQ ID NO: 59:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 84 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: peptide
/ ORIGINAL SOURCE:
/ ORGANISM: Ancylostoma caninum
/
/ US-08-486-399-59
/
/
/ Query Match 25.6%; Score 100; DB 2; Length 84;
/ Best Local Similarity 30.4%; Pred. No. 0.0012;
/ Matches 21; Conservative 8; Mismatches 24; Indels 16; Gaps 3;
/
/ QY 11 KCPSEIFRCDGR-CQRF-----PNVVKPLCIKICAPGCVCRIGYLRNKK 57
/ Db 5 QCGENEKYDSGSKCKKCYDGVVEEDDEPNV---PCLVRVCHQDCVCEGFRNKD 61
/
/ QY 58 KVCVPRSKC 66
/ Db 62 DKCVSAEDC 70
/
/ RESULT 9
/ US-08-634-641-59
/ Sequence 59, Application US/08634641
/ Patent No. 5955294
/ GENERAL INFORMATION:
/ APPLICANT: Vlasuk, George P. Vlasuk
/ APPLICANT: Stanssens, Patrick Eric Hugo
/ APPLICANT: Mensens, Joris Hilda Lieven
/ APPLICANT: Lauwereys, Marc Josef
/ APPLICANT: Laroche, Yves Rene
/ APPLICANT: Jespers, Laurent Stephane
/ APPLICANT: Gansmans, Yannick Georges Jozef
/ APPLICANT: Moyle, Matthew
/ APPLICANT: Bergum, Peter W.
/ TITLE OF INVENTION: NEMATODE-EXTRACTED ANTICOAGULANT
/ NUMBER OF SEQUENCES: 356
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Lyon & Lyon
/ STREET: 633 West Fifth Street
/ CITY: Suite 4700
/ STATE: Los Angeles
/ COUNTRY: California
/ COUNTRY: U.S.A.
/ ZIP: 90071
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
/ MEDIUM TYPE: storage
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: IBM P.C. DOS 5.0
/ SOFTWARE: Word Perfect 5.1
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/634,641
/ FILING DATE: April 19, 1996
/ CLASSIFICATION: 530
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: PCT/US95/13231
```

```
/
/
/ APPLICATION NUMBER: US/08/461,965
/ FILING DATE: June 5, 1995
/ CLASSIFICATION: 530
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/326,110
/ FILING DATE: October 18, 1994
/ ATTORNEY/AGENT INFORMATION:
/ NAME: BIGGS, SUZANNE L.
/ REGISTRATION NUMBER: 30,158
/ REFERENCE/DOCKET NUMBER: 210/243
/ TELEPHONE: (213) 489-1600
/ TELEFAX: (213) 955-0440
/ TELEX: 67-3510
/ INFORMATION FOR SEQ ID NO: 59:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 84 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: peptide
/ ORIGINAL SOURCE:
/ ORGANISM: Ancylostoma caninum
/
/ US-08-461-965-59
/
/
/ Query Match 25.6%; Score 100; DB 2; Length 84;
/ Best Local Similarity 30.4%; Pred. No. 0.0012;
/ Matches 21; Conservative 8; Mismatches 24; Indels 16; Gaps 3;
/
/ QY 11 KCPSEIFRCDGR-CQRF-----PNVVKPLCIKICAPGCVCRIGYLRNKK 57
/ Db 5 QCGENEKYDSGSKCKKCYDGVVEEDDEPNV---PCLVRVCHQDCVCEGFRNKD 61
/
/ QY 58 KVCVPRSKC 66
/ Db 62 DKCVSAEDC 70
/
/ RESULT 9
/ US-08-634-641-59
/ Sequence 59, Application US/08634641
/ Patent No. 5955294
/ GENERAL INFORMATION:
/ APPLICANT: Vlasuk, George P. Vlasuk
/ APPLICANT: Stanssens, Patrick Eric Hugo
/ APPLICANT: Mensens, Joris Hilda Lieven
/ APPLICANT: Lauwereys, Marc Josef
/ APPLICANT: Laroche, Yves Rene
/ APPLICANT: Jespers, Laurent Stephane
/ APPLICANT: Gansmans, Yannick Georges Jozef
/ APPLICANT: Moyle, Matthew
/ APPLICANT: Bergum, Peter W.
/ TITLE OF INVENTION: NEMATODE-EXTRACTED ANTICOAGULANT
/ NUMBER OF SEQUENCES: 356
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Lyon & Lyon
/ STREET: 633 West Fifth Street
/ CITY: Suite 4700
/ STATE: Los Angeles
/ COUNTRY: California
/ COUNTRY: U.S.A.
/ ZIP: 90071
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
/ MEDIUM TYPE: storage
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: IBM P.C. DOS 5.0
/ SOFTWARE: Word Perfect 5.1
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/634,641
/ FILING DATE: April 19, 1996
/ CLASSIFICATION: 530
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: PCT/US95/13231
```



```

; FILING DATE: October 17, 1995
; APPLICATION NUMBER: 08/486,399
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/486,397
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/465,380
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/461,965
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/326,110
; FILING DATE: October 18, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BIGGS, SUZANNE L.
; REGISTRATION NUMBER: 30,158
; REFERENCE/DOCKET NUMBER: 219/136
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 84 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Ancylostoma caninum
; US-08-634-641-59

Query Match 25.6%; Score 100; DB 2; Length 84;
Best Local Similarity 30.4%; Pred. No. 0.0012;
Matches 21; Conservative 8; Mismatches 24; Indels 16; Gaps 3;

Qy 11 KCPSEIFSRCDGR-CORFC-----PNVVKPLCIKICAPGCVCLGYLNKK 57
Db 5 QCGENEKIDSGSKCEDKCKYDGVBEEDDEPNV---PCLVRVCHQDCVCEBGFYRNKD 61

Qy 58 KVCVPRSKC 66
Db 62 DKCVSAEDC 70

RESULT 10
US-09-249-471-59
; Sequence 59, Application US/09249471
; Patent No. 6040441
; GENERAL INFORMATION:
; APPLICANT: Vlasuk, George Phillip
; APPLICANT: Stanssens, Patrick Eric Hugo
; APPLICANT: Messens, Joris Hilda Lieven
; APPLICANT: Lauwereys, Marc Josef
; APPLICANT: Laroche, Yves Rene
; APPLICANT: Jespers, Laurent Stephane
; APPLICANT: Ganssemans, Yannick Georges Jozef
; APPLICANT: Moyle, Matthew
; APPLICANT: Bergum, Peter W.
; TITLE OF INVENTION: NEMATODE-EXTRACTED SERINE PROTEASE
; TITLE OF INVENTION: INHIBITORS AND ANTICOAGULANT
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 356
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
```

```

; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/249,471
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/809,455
; FILING DATE: April 17, 1997
; APPLICATION NUMBER: PCT/US95/13231
; FILING DATE: October 17, 1995
; APPLICATION NUMBER: 08/486,399
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/486,397
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/465,380
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/461,965
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/326,110
; FILING DATE: October 18, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BIGGS, SUZANNE L.
; REGISTRATION NUMBER: 30,158
; REFERENCE/DOCKET NUMBER: 216/270
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 84 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Ancylostoma caninum
; US-09-249-471-59

Query Match 25.6%; Score 100; DB 3; Length 84;
Best Local Similarity 30.4%; Pred. No. 0.0012;
Matches 21; Conservative 8; Mismatches 24; Indels 16; Gaps 3;

Qy 11 KCPSEIFSRCDGR-CORFC-----PNVVKPLCIKICAPGCVCLGYLNKK 57
Db 5 QCGENEKIDSGSKCEDKCKYDGVBEEDDEPNV---PCLVRVCHQDCVCEBGFYRNKD 61

Qy 58 KVCVPRSKC 66
Db 62 DKCVSAEDC 70

RESULT 11
US-09-249-472-59
; Sequence 59, Application US/09249472
; Patent No. 6046318
; GENERAL INFORMATION:
; APPLICANT: Vlasuk, George Phillip
; APPLICANT: Stanssens, Patrick Eric Hugo
; APPLICANT: Messens, Joris Hilda Lieven
; APPLICANT: Lauwereys, Marc Josef
; APPLICANT: Laroche, Yves Rene
; APPLICANT: Jespers, Laurent Stephane
; APPLICANT: Ganssemans, Yannick Georges Jozef
; APPLICANT: Moyle, Matthew
; APPLICANT: Bergum, Peter W.
; TITLE OF INVENTION: NEMATODE-EXTRACTED SERINE PROTEASE
; TITLE OF INVENTION: INHIBITORS AND ANTICOAGULANT
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 356
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles
```

STATE: California
COUNTRY: U.S.A.
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/249,472
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/809,455
FILING DATE: April 17, 1997
APPLICATION NUMBER: PCT/US95/13231
FILING DATE: October 17, 1995
APPLICATION NUMBER: 08/486,399
FILING DATE: June 5, 1995
APPLICATION NUMBER: 08/486,397
FILING DATE: June 5, 1995
APPLICATION NUMBER: 08/465,380
FILING DATE: June 5, 1995
APPLICATION NUMBER: 08/461,965
FILING DATE: June 5, 1995
APPLICATION NUMBER: 08/326,110
FILING DATE: October 18, 1994
ATTORNEY/AGENT INFORMATION:
NAME: BIGGS, SUZANNE L.
REGISTRATION NUMBER: 30,158
REFERENCE/DOCKET NUMBER: 216/270
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510

INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 84 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Ancylostoma caninum
US-09-249-472-59

Query Match 25.6%; Score 100; DB 3; Length 84;
Best Local Similarity 30.4%; Pred. No. 0.0012;
Matches 21; Conservative 8; Mismatches 24; Indels 16; Gaps 3;

QY 11 KCPSEIFSRCDGR-CQRF-----PNVVPKPLCIKICAPGCVCLGYLRNKK 57
Db 5 QCGENEKYDSCGSKCKCKYDGVVEEDDEPNV---PCLVRVCHQDCVCEGFRNKD 61

QY 58 KVCVPRSKC 66
Db 62 DKCVSAEDC 70

RESULT 12

US-09-249-451-59
Sequence 59, Application US/09249451
Patent No. 6087487

GENERAL INFORMATION:

APPLICANT: Vlasuk, George Phillip
APPLICANT: Stanssens, Patrick Eric Hugo
APPLICANT: Messens, Joris Hilda Lieven
APPLICANT: Lauwereys, Marc Josef
APPLICANT: Larocche, Yves Rene
APPLICANT: Jespers, Laurent Stephane
APPLICANT: Ganssemans, Yannick Georges Jozef
APPLICANT: Moyle, Matthew
APPLICANT: Bergum, Peter W.
TITLE OF INVENTION: NEMATODE-EXTRACTED SERINE PROTEASE

TITLE OF INVENTION: INHIBITORS AND ANTICOAGULANT
TITLE OF INVENTION: PROTEIN
NUMBER OF SEQUENCES: 356
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/249,451
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/809,455
FILING DATE: April 17, 1997
APPLICATION NUMBER: PCT/US95/13231
FILING DATE: October 17, 1995
APPLICATION NUMBER: 08/486,399
FILING DATE: June 5, 1995
APPLICATION NUMBER: 08/486,397
FILING DATE: June 5, 1995
APPLICATION NUMBER: 08/465,380
FILING DATE: June 5, 1995
APPLICATION NUMBER: 08/461,965
FILING DATE: June 5, 1995
APPLICATION NUMBER: 08/326,110
FILING DATE: October 18, 1994
ATTORNEY/AGENT INFORMATION:
NAME: BIGGS, SUZANNE L.
REGISTRATION NUMBER: 30,158
REFERENCE/DOCKET NUMBER: 216/270
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510

INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 84 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Ancylostoma caninum
US-09-249-451-59

Query Match 25.6%; Score 100; DB 3; Length 84;
Best Local Similarity 30.4%; Pred. No. 0.0012;
Matches 21; Conservative 8; Mismatches 24; Indels 16; Gaps 3;

QY 11 KCPSEIFSRCDGR-CQRF-----PNVVPKPLCIKICAPGCVCLGYLRNKK 57
Db 5 QCGENEKYDSCGSKCKCKYDGVVEEDDEPNV---PCLVRVCHQDCVCEGFRNKD 61

QY 58 KVCVPRSKC 66
Db 62 DKCVSAEDC 70

RESULT 13

US-08-809-455-59
Sequence 59, Application US/08809455
Patent No. 6090916
GENERAL INFORMATION:
APPLICANT: Vlasuk, George Phillip
APPLICANT: Stanssens, Patrick Eric Hugo

APPLICANT: Messens, Joris Hilda Lieven
APPLICANT: Lauwereys, Marc Josef
APPLICANT: Laroche, Yves Rene
APPLICANT: Jespers, Laurent Stephane
APPLICANT: Gansemans, Yannick Georges Jozef
APPLICANT: Moyle, Matthew
APPLICANT: Bergum, Peter W.
TITLE OF INVENTION: NEMATODE-EXTRACTED SERINE PROTEASE
TITLE OF INVENTION: INHIBITORS AND ANTICOAGULANT
TITLE OF INVENTION: PROTEIN
NUMBER OF SEQUENCES: 356
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/809,455
FILING DATE: April 17, 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/13231
FILING DATE: October 17, 1995
APPLICATION NUMBER: 08/486,399
FILING DATE: June 5, 1995
APPLICATION NUMBER: 08/465,380
FILING DATE: June 5, 1995
APPLICATION NUMBER: 08/461,965
FILING DATE: June 5, 1995
APPLICATION NUMBER: 08/326,110
FILING DATE: October 18, 1994
ATTORNEY/AGENT INFORMATION:
NAME: BIGGS, SUZANNE L.
REGISTRATION NUMBER: 30,158
REFERENCE/DOCKET NUMBER: 216/270
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 84 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Ancylostoma caninum
US-08-809-455-59

Query Match 25.6%; Score 100; DB 3; Length 84;
Best Local Similarity 30.4%; Pred. No. 0.0012;
Matches 21; Conservative 8; Mismatches 24; Indels 16; Gaps 3;
QY 11 KCPSEIFSRCDGR-CORFC-----PNNVFKPLCIKICAPGCVCRGLYLNKK 57
Db 5 QCGENKIDCGSKCEDKCKYGVVEEDDEEPNV---PCLVRVCHQDCVCEGFRNKK 61

QY 58 KVCVPSK 66
Db 62 DKCVSAEDC 70

RESULT 14

US-09-249-461-59
Sequence 59, Application US/09249461
Patent No. 6096877
GENERAL INFORMATION:
APPLICANT: Vlasuk, George Phillip
APPLICANT: Stanssens, Patrick Eric Hugo
APPLICANT: Messens, Joris Hilda Lieven
APPLICANT: Lauwereys, Marc Josef
APPLICANT: Laroche, Yves Rene
APPLICANT: Jespers, Laurent Stephane
APPLICANT: Gansemans, Yannick Georges Jozef
APPLICANT: Moyle, Matthew
APPLICANT: Bergum, Peter W.
TITLE OF INVENTION: NEMATODE-EXTRACTED SERINE PROTEASE
TITLE OF INVENTION: INHIBITORS AND ANTICOAGULANT
TITLE OF INVENTION: PROTEIN
NUMBER OF SEQUENCES: 356
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/249,461
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/809,455
FILING DATE: April 17, 1997
APPLICATION NUMBER: PCT/US95/13231
FILING DATE: October 17, 1995
APPLICATION NUMBER: 08/486,399
FILING DATE: June 5, 1995
APPLICATION NUMBER: 08/486,397
FILING DATE: June 5, 1995
APPLICATION NUMBER: 08/465,380
FILING DATE: June 5, 1995
APPLICATION NUMBER: 08/461,965
FILING DATE: June 5, 1995
APPLICATION NUMBER: 08/326,110
FILING DATE: October 18, 1994
ATTORNEY/AGENT INFORMATION:
NAME: BIGGS, SUZANNE L.
REGISTRATION NUMBER: 30,158
REFERENCE/DOCKET NUMBER: 216/270
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 84 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Ancylostoma caninum
US-09-249-461-59

Query Match 25.6%; Score 100; DB 3; Length 84;
Best Local Similarity 30.4%; Pred. No. 0.0012;
Matches 21; Conservative 8; Mismatches 24; Indels 16; Gaps 3;
QY 11 KCPSEIFSRCDGR-CORFC-----PNNVFKPLCIKICAPGCVCRGLYLNKK 57
Db 5 QCGENKIDCGSKCEDKCKYGVVEEDDEEPNV---PCLVRVCHQDCVCEGFRNKK 61

Db 5 QCGENEKYDCGSKCKKCYDGVVEEDDEPNV---PCLVRVCHQDCVCBEGFYRNKD 61

QY 58 KVCVPRSKC 66
Db 62 DKCVSAEDC 70

RESULT 15

US-09-249-448-59
; Sequence 59, Application US/09249448
; Patent No. 6121435

; GENERAL INFORMATION:

; APPLICANT: Vlasuk, George Phillip
; APPLICANT: Stanssens, Patrick Eric Hugo
; APPLICANT: Messens, Joris Hilda Lieven
; APPLICANT: Lauwereys, Marc Josef
; APPLICANT: Laroche, Yves Rene
; APPLICANT: Jespers, Laurent Stephane
; APPLICANT: Gansemans, Yannick Georges Jozef
; APPLICANT: Moyle, Matthew
; APPLICANT: Bergum, Peter W.

; TITLE OF INVENTION: NEMATODE-EXTRACTED SERINE PROTEASE
; TITLE OF INVENTION: INHIBITORS AND ANTICOAGULANT
; NUMBER OF SEQUENCES: 356

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/249,448

; FILING DATE:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/809,455
; FILING DATE: April 17, 1997
; APPLICATION NUMBER: PCT/US95/13231
; FILING DATE: October 17, 1995
; APPLICATION NUMBER: 08/486,399
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/486,397
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/465,380
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/461,965
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/326,110
; FILING DATE: October 18, 1994

; ATTORNEY/AGENT INFORMATION:

; NAME: BIGGS, SUZANNE L.
; REGISTRATION NUMBER: 30,158
; REFERENCE/DOCKET NUMBER: 216/270
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510

; INFORMATION FOR SEQ ID NO: 59:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 84 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; ORIGINAL SOURCE:

; ORGANISM: Ancylostoma caninum

US-09-249-448-59

Query Match 25.6%; Score 100; DB 3; Length 84;
Best Local Similarity 30.4%; Pred. No. 0.0012;
Matches 21; Conservative 8; Mismatches 24; Indels 16; Gaps 3;

QY 11 KCPSNEIFSRCDGR-CQRF-----PNTVPKPLCIKICAPGCVCLGVLRNKK 57
Db 5 QCGENEKYDCGSKCKKCYDGVVEEDDEPNV---PCLVRVCHQDCVCBEGFYRNKD 61
QY 58 KVCVPRSKC 66
Db 62 DKCVSAEDC 70

Search completed: November 17, 2003, 08:10:25
Job time : 23 secs

Result No.	Score	Query			DB	ID	Description
		Match	Length				
1	80.5	20.6	63	2	S07127	chymotrypsin/elast	
2	77.5	19.8	63	2	S08572	chymotrypsin/elast	
3	70.5	18.0	62	2	S35098	trypsin inhibitor	
4	69	17.6	60	2	S31723	metallothionein -	
5	69	17.6	60	2	B27490	metallothionein B	
6	68	17.4	60	2	S30567	metallothionein -	
7	67	17.1	62	2	A53640	metallothionein 4	
8	65	16.6	62	2	B53640	metallothionein IV	
9	62	15.9	61	2	A37425	metallothionein 2	
10	62	15.9	62	2	S54335	metallothionein 2c	
11	62	15.9	62	2	S54336	metallothionein-2a	
12	61.5	15.7	61	2	A23889	metallothionein 1	
13	60.5	15.5	61	1	SMH1A	metallothionein 1A	
14	60.5	15.5	61	2	JCI449	metallothionein A	
15	60	15.3	52	2	S65712	metallothionein 1	
16	59.5	15.2	61	1	SMH1B	metallothionein 1B	
17	59.5	15.2	61	2	S69277	metallothionein 1R	
18	59.5	15.2	63	2	S08190	metallothionein 1	
19	59.5	15.2	63	2	A34905	metallothionein 1	
20	59	15.1	60	2	JC2420	metallothionein -	
21	59	15.1	61	1	SMBO2	metallothionein II	
22	59	15.1	61	2	S00808	metallothionein Ia	
23	59	15.1	61	2	S00809	metallothionein Ib	
24	59	15.1	65	2	A38739	metallothionein -	
25	58.5	15.0	61	1	SMMSI	metallothionein I	
26	58.5	15.0	61	2	S47651	metallothionein IH	
27	58.5	15.0	61	2	I54574	metallothionein-1	
28	58	14.8	60	2	S00810	metallothionein Ic	
29	57.5	14.7	60	1	SMH1A	metallothionein 1A	

Oy 9 RGKCPSEIIFSRCDGRGCRFCFNVVPKPICIKICAPGCVC - LGYLNNKKKVCVPRSKC 66
| | | | | : | : | : | : | : | : | : | : | :
Ddb 1 RKPCGNEVWTEGTG-CELCKGODENTPCALMCRPPSCSGRGMERTHDGKCVPVSEC 59
| | | | | : | : | : | : | : | : | : | : | :

RESULT 3

S35098
trypsin inhibitor - pig roundworm
C;Species: Ascaris suum (pig roundworm)
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 31-Oct-1997
C;Accession: S35098
R;Gronenborn, A.M.; Nilges, M.; Peanasky, R.J.; Clore, G.M.
Biochemistry 29, 183-189, 1990
A;Title: Sequential resonance assignment and secondary structure determination of the A8
A;Reference number: S35098; MUID:90212567; PMID:2322539
A;Accession: S35098
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-62 <GRO>
C;Superfamily: roundworm trypsin inhibitor

Query Match 18.0%; Score 70.5; DB 2; Length 62;
Best Local Similarity 29.5%; Pred. No. 2.2;
Matches 18; Conservative 12; Mismatches 22; Indels 9; Gaps 5;

QY 11 KCPSP-NEIFSRCDGRQRCF-PNVVPKPLCIKICAP---GCVCRLGYLRNKKVKVPSRX 65

DB 4 KCTKPNEQWTKCGG-CEGTCAQKIVP---CTRECKPRCEIASAGFVRDAQNCIKFED 59

QY 66 C 66

DB 60 C 60

RESULT 4

S31723
metallothionein - northern pike
C;Species: Esox lucius (northern pike)
C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 20-Aug-1999
C;Accession: S38334; S17175; S15503; S31723
R;Kille, P.; Kay, J.; Sweeney, G.E.
Biochim. Biophys. Acta 1216, 55-64, 1993
A;Title: Analysis of regulatory elements flanking metallothionein genes in Cd-tolerant B
A;Reference number: S38334; MUID:94032489; PMID:8218416
A;Accession: S38334
A;Molecule type: DNA
A;Residues: 1-60 <KT3>
A;Cross-references: EMBL:X70042; NID:g62782; PIDN:CAA49636.1; PID:g62783
A;Note: the authors translated the codon ACT for residue 9 as Ser
R;Kille, P.; Stephens, P.E.; Kay, J.
Biochim. Biophys. Acta 1089, 407-410, 1991
A;Title: Elucidation of cDNA sequences for metallothioneins from rainbow trout, stone l
A;Reference number: S16996; MUID:91316146; PMID:1859844
A;Accession: S17175
A;Molecule type: mRNA
A;Residues: 1-60 <KIL>
A;Cross-references: EMBL:X59392; NID:g62780; PIDN:CAA42035.1; PID:g62781
C;Genetics:
C;Introns: 9/1; 31/1
C;Superfamily: metallothionein
C;Keywords: chelation; metal binding; metal-thiolate cluster

Query Match 17.6%; Score 69; DB 2; Length 60;
Best Local Similarity 35.2%; Pred. No. 3;
Matches 19; Conservative 3; Mismatches 18; Indels 14; Gaps 3;

QY 7 GGRGKCPSPNEIFSRCDGRQRCF-PNVVPKPLCIKICAPGCVRLGYLRNKKVKC 60

DB 15 GGSCKC-SNCACTSKCKSCCPSPGCSK-----CASGCIC-----KGKTC 54

RESULT 5.

B27490
metallothionein B - rainbow trout
C;Species: Oryzias latipes mykiss (rainbow trout)
C;Date: 18-Jul-1988 #sequence_revision 18-Jul-1988 #text_change 20-Aug-1999

C;Accession: A30818; B27490; S16997; S15509
R;Zafarullah, M.; Bonham, K.; Gedamu, L.
Mol. Cell. Biol. 8, 4469-4476, 1988
A;Title: Structure of the rainbow trout metallothionein B gene and characterization <
A;Reference number: A30818; MUID:89039876; PMID:3185557
A;Accession: A30818
A;Molecule type: DNA
A;Residues: 1-60 <ZAF>
A;Cross-references: GB:M18104; NID:g213816; PIDN:AAA49566.1; PID:g213817
R;Bonham, K.; Zafarullah, M.; Gedamu, L.
DNA 6, 519-528, 1987
A;Title: The rainbow trout metallothioneins: molecular cloning and characterization <
A;Reference number: A27490; MUID:88111026; PMID:2448099
A;Accession: B27490
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-60 <BON>
A;Cross-references: GB:M18104; NID:g213816; PIDN:AAA49566.1; PID:g213817
R;Kille, P.; Stephens, P.E.; Kay, J.
Biochim. Biophys. Acta 1089, 407-410, 1991
A;Title: Elucidation of cDNA sequences for metallothioneins from rainbow trout, stone
A;Reference number: S16996; MUID:91316146; PMID:1859844
A;Accession: S16997
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-60 <KIL>
A;Cross-references: EMBL:X59394; NID:g64332; PIDN:CAA42037.1; PID:g64333
C;Superfamily: metallothionein

Query Match 17.6%; Score 69; DB 2; Length 60;
Best Local Similarity 37.0%; Pred. No. 3;
Matches 20; Conservative 2; Mismatches 18; Indels 14; Gaps 3;

QY 7 GGRGKCPSPNEIFSRCDGRQRCF-PNVVPKPLCIKICAPGCVRLGYLRNKKVKC 60

DB 15 GGSCKC-SNCACTSKCKSCCPSPGCSK-----CASGCVC-----KGKTC 54

RESULT 6

S30567
metallothionein - plaice
C;Species: Pleuronectes platessa (plaice)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 20-Aug-1999
C;Accession: S30567
R;Leaver, M.J.; George, S.G.
submitted to the EMBL Data Library, November 1990
A;Reference number: S30567
A;Accession: S30567
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-60 <LEA>
A;Cross-references: EMBL:X56743; NID:g64237; PIDN:CAA40067.1; PID:g64238
C;Superfamily: metallothionein

Query Match 17.4%; Score 68; DB 2; Length 60;
Best Local Similarity 35.2%; Pred. No. 3.7;
Matches 19; Conservative 3; Mismatches 18; Indels 14; Gaps 3;

QY 7 GGRGKCPSPNEIFSRCDGRQRCF-PNVVPKPLCIKICAPGCVRLGYLRNKKVKC 60

DB 15 GGSCTC-KNCSCTTCNKSCCPSPGCPK-----CASGCVC-----KGKTC 54

RESULT 7

A53640
N;Alternate names: metallothionein IV
C;Species: Homo sapiens (man)
C;Date: 05-Jan-1996 #sequence_revision 05-Jan-1996 #text_change 03-Dec-1999
C;Accession: A53640
R;Quaife, C.J.; Findley, S.D.; Erickson, J.C.; Froelick, G.J.; Kelly, E.J.; Zambrowic
Biochemistry 33, 7250-7259, 1994
A;Title: Induction of a new metallothionein isoform (MT-IV) occurs during differenti

10 GGSTCTAGSCKCKA---CRPS-CKGSCCSCCPVG-CAK-CAQGCVC 51

RESULT 13

```

SMWHUIA
metallothionein 1A - human
C:Species: Homo sapiens (man)
C:Date: 24-Jun-1987 #sequence_revision 09-Sep-1994 #text_change 22-Jun-1999
C:Accession: A24502
R:Richards, R.I.; Heguy, A.; Karin, M.
Cell 37, 263-272, 1984
A:Title: Structural and functional analysis of the human metallothionein-IA gene: differ
A:Reference number: A24502; MUID:84205649; PMID:6327055
A:Accession: A24502
A:Molecule type: DNA
A:Residues: 1-61 <R1C>
A:Cross-references: GB:K01383; NID:g187536; PIDN:AAA59586.1; PID:G386864
C:Genetics:
A:Gene: GDB:MT1A; MT1
A:Cross-references: GDB:125559; OMIM:156350
A:Map position: 16q13-16q13
A:Introns: 10/1; 32/1
C:Superfamily: metallothionein
C:Keywords: acetylated amino end; chelation; metal binding; metal-thiolate cluster
F:1-23/Domain: beta <NH2>
F:30-61/Domain: alpha <ALP>
F:1/Modified site: acetylated amino end (Met) #status predicted
F:5/7,13,15,19,21,24,26,29/Binding site: transition metal ions (Cys) #status predicted
F:33,34,36,37,41,44,48,50,57,59,60/Binding site: transition metal ions (Cys) #status pre
Query Match 15.5%; Score 60.5; DB 1; Length 61;
Best Local Similarity 35.0%; Pred. No. 21;
Matches 14; Conservative 7; Mismatches 14; Indels 5; Gaps 4;
QY 10 GKCPNFIKRCRGRFCFNNVVKPLKIKICAPGCVCVR 49
Db 17 GSKCKRE--CKNGS--CKKSCSCCPMS--CAK--CAOGCICK 51

```

RESULT 14

JC1449 metallothionein A - rainbow trout
C:Species: Oncorhynchus mykiss (rainbow trout)
C:Date: 30-Sep-1993 #sequence revision 30-Sep-1993 #text_change 20-Aug-1999
C:Accession: JC1449; A27490; S16996; S15508
R:Hong, Y.; Schartl, M.
Gene 120, 277-279, 1992
A:Title: Structure of the rainbow trout metallothionein A gene.
A:Reference number: JC1449; MUID:93013046; PMID:1398141
A:Accession: JC1449
A:Molecule type: DNA
A:Residues: 1-61 <ON>
A:Cross-references: GB:M81800; NID:g213812; PIDN:AAA49564.1; PID:g213813
R:Bonham, K.; Zafarullah, M.; Gedamu, L.
DNA 6, 519-528, 1987
A:Title: The rainbow trout metallothioneins: molecular cloning and characterization of b
A:Reference number: A27490; MUID:88111026; PMID:2448099
A:Accession: A27490
A:Molecule type: mRNA
A:Residues: 1-61 <ON>
A:Cross-references: GB:M81803; NID:g213814; PIDN:AAA49565.1; PID:g213815
R:Kille, P.; Stephens, P.E.; Kay, J.
Biochim. Biophys. Acta 1089, 407-410, 1991
A:Title: Elucidation of cDNA sequences for metallothioneins from rainbow trout, stone lo
A:Reference number: S16996; MUID:91316146; PMID:1859844
A:Accession: S16996
A:Molecule type: mRNA
A:Residues: 1-61 <KL>
A:Cross-references: EMBL:X59395; NID:g64330; PIDN:CAA42038.1; PID:g64331
C:Genetics:
A:Gene: tMT-A
A:Introns: 9/-; 32/1

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2003, 08:10:28 ; Search time 11 Seconds
(without alignments)
286.435 Million cell updates/sec

Title: US-09-506-978-1
Perfect score: 391
Sequence: 1 GGGGGLGGRKCPSPNEIFSR.....CRLGLYLRNKKKVCVPRSKG 67

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 6867

Minimum DB seq length: 0
Maximum DB seq length: 67

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	104.5	26.7	56	1 AMCI_APIME	P56682 apis mellif
2	80.5	20.6	63	1 ICE1_ASCSU	P07851 ascaris suu
3	78.5	20.1	65	1 ICE2_ASCSU	P07852 ascaris suu
4	73	18.7	60	1 MT_PERFL	P52725 perca fluvi
5	70.5	18.0	62	1 ITR1_ASCSU	P19398 ascaris suu
6	69	17.6	60	1 MTB_ONCMY	P09862 oncorhynch
7	69	17.6	60	1 MT_ESOLU	P25127 esox lucius
8	68	17.4	60	1 MT_PLEPL	P07216 pleuronecte
9	68	17.4	60	1 MT_PSEAM	P55945 pseudopleur
10	67	17.1	62	1 MT4_HUMAN	P47944 homo sapien
11	66.5	17.0	61	1 MT_BALMY	O18942 balaena mys
12	66	16.9	60	1 MTB_SALSA	P52720 salmo salar
13	65	16.6	60	1 MT_ZOAVI	P52728 zoarces viv
14	65	16.6	62	1 MT4_MOUSE	P47945 mus musculus
15	63.5	16.2	60	1 MT_AMBME	O42152 ambystoma m
16	62.5	16.0	62	1 MT4_CANFA	O9tuis canis famil
17	62	15.9	62	1 MT2A_RABIT	P18055 oryctolagus
18	62	15.9	62	1 MT2C_RABIT	P80290 oryctolagus
19	61.5	15.7	61	1 MT1_BOVIN	P04731 homo sapien
20	60.5	15.5	61	1 MT1A_HUMAN	P14425 stenella co
21	60.5	15.5	61	1 MT2_STECO	P09861 oncorhynch
22	60.5	15.5	61	1 MT2_ONCMY	O9ib50 pagrus majo
23	60	15.3	60	1 MTB_FAGMA	P07438 homo sapien
24	59.5	15.2	61	1 MT1B_HUMAN	Q93083 homo sapien
25	59.5	15.2	61	1 MT1B_HUMAN	Q9xst5 canis famil
26	59.5	15.2	61	1 MT2_CANFA	P17512 caenorhabdi
27	59.5	15.2	62	1 MT2_CABEL	P15786 columba liv
28	59.5	15.2	63	1 MT1_COLLI	P52724 chaenocepha
29	59	15.1	60	1 MTB_CHAAC	O13259 chionodrac
30	59	15.1	60	1 MTB_CHIHA	Q9pt59 dicentrach
31	59	15.1	60	1 MTB_DICLA	Q92145 pagothenia
32	59	15.1	60	1 MTB_PAGBE	O13257 lixa aurata
33	59	15.1	60	1 MT_LIZAU	

ALIGNMENTS

RESULT 1

```
AMCI_APIME          STANDARD;          PRT;          56 AA.
ID AC P56682;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DB Chymotrypsin inhibitor (AMCI).
OS Apis mellifera (Honeybee).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea;
OC Apidae; Apis.
OX NCBI_TaxID=7460;
RN [1]
RP SEQUENCE, AND STRUCTURE BY NMR.
RC TISSUE=Hemolymph; PubMed=10411628;
RX MEDLINE=99339935;
RA Bania J., Stachowiak D., Polanowski A.;
RT "Primary structure and properties of the cathepsin G/chymotrypsin
RL inhibitor from the larval hemolymph of Apis mellifera.";
RL Eur. J. Biochem. 262:680-687(1999).
CC -!- FUNCTION: Chymotrypsin and cathepsin G inhibitor.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Contains 1 TIL (trypsin inhibitory-like) domain.
DR PDB; ICCT; 12-MAR-99.
DR InterPro; IPR002919; TIL_Cysrich.
DR Pfam; PF01826; TIL; 1.
KW Serine protease inhibitor; 3D-structure.
FT DOMAIN 3 56 TIL.
FT DISULFID 3 36
FT DISULFID 12 32
FT DISULFID 16 28
FT DISULFID 20 56
FT DISULFID 38 50
FT TURN 5 6
FT STRAND 7 13
FT TURN 15 16
FT STRAND 33 38
FT TURN 40 41
FT STRAND 43 45
FT TURN 46 48
FT STRAND 49 51
FT HELIX 53 56
SQ SEQUENCE 56 AA; 5973 MW; 092B2815AE6B2B7F CRC64;
```

Query Match 26.7%; Score 104.5; DB 1; Length 56;

Best Local Similarity 36.8%; Pred. No. 5.8e-05;

Matches 21; Conservative 7; Mismatches 26; Indels 3; Gaps 2;

QY 11 KCPSPNEIFSRCDGRQCRFCFNVVFKP-LCTIKICAPGVCRLGYLRNKKKVCVPRSKC 66

Db 2 ECGPNEVENTCGSACAPTCAQ--PKTRICTMQRCQCGFLRNGEGACVLPPENC 56

RESULT 2

ICE1_ASCSU

```
ID ICE1_ASCSU STANDARD; PRT; 63 AA.
AC P07851; 077419;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DE 15-SEP-2003 (Rel. 42, Last annotation update)
DE Chymotrypsin/elastase isoform 1 (C/E-1 inhibitor) (Asc/E-1).
OS Ascaris suum (Pig roundworm) (Ascaris lumbricoides).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Ascarididae; Ascaridoidea;
OC Ascarididae; Ascaris.
OX NCBI_TaxID=6253;
RN [1]
RP SEQUENCE.
RX MEDLINE=84255715; PubMed=6564898;
RA Babin D.R., Peanasky R.J., Goos S.M.;
RT "The isoforms of chymotrypsin/elastase from Ascaris
lumbicoides: the primary structure."
RL Arch. Biochem. Biophys. 232:143-161 (1984).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98297373; PubMed=9635450;
RA Lu C.C., Nguyen T., Morris S., Hill D., Sakanari J.A.;
RT "Anisakis simplex: mutational bursts in the reactive site centers of
serine protease inhibitors from an ascarid nematode."
RL Exp. Parasitol. 89:257-261 (1998).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF COMPLEX WITH ELASTASE.
RX MEDLINE=95006335; PubMed=7922044;
RA Huang K., Strynadka N.C., Bernard V.D., Peanasky R.J., James M.N.;
RT "The molecular structure of the complex of Ascaris
chymotrypsin/elastase inhibitor with porcine elastase."
RL Structure 2:679-689 (1994).
CC -!- FUNCTION: Defend the organism against the host's proteinases.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Contains 1 TIL (Trypsin inhibitory-like) domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U94499; AAC61300.1; -.
DR PDB; 1EAI; 05-APR-99.
DR InterPro: IPR002919; TIL_Cysrich.
DR Pfam; PF01826; TIL; 1.
KW Serine protease inhibitor; 3D-structure.
FT DOMAIN 5 60
FT DISULFID 5 38 TIL.
FT DISULFID 14 33
FT DISULFID 17 29
FT DISULFID 21 60
FT DISULFID 40 54
FT ACT_SITE 31 32
FT CONFLICT 4 4
FT CONFLICT 23 24
FT TURN 3 4
FT TURN 7 8
FT STRAND 10 12
FT STRAND 15 15
FT STRAND 20 20
FT TURN 25 26
FT STRAND 28 30
FT STRAND 34 34
FT STRAND 37 39
FT HELIX 42 44
FT TURN 45 45
FT STRAND 46 48
FT TURN 50 51
FT STRAND 54 56
FT HELIX 57 59
SQ SEQUENCE .. 63 AA; 6862 MW; 5DC10DE75B375F16 CRC64;
```

```
Query Match 20.6%; Score 80.5; DB 1; Length 63;
Best Local Similarity 30.3%; Pred. No. 0.024;
Matches 20; Conservative 10; Mismatches 23; Indels 13; Gaps 4;
QY 8 GRGKPSNFIKRCDCG---RC-----QRFQPVVVKPLCKIKICAFGCVCRGLGYLRNKKVC 60
DB 1 QGESCQPNVWTECTGCEMKCGPDENPTCLMCRPFSC--ECSFG---RGMRTNDGKC 54
QY 61 VPRSKC 66
DB 55 IPASQC 60
```

RESULT 3

```
ICE2_ASCSU STANDARD; PRT; 65 AA.
AC P07852;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DE 15-SEP-2003 (Rel. 42, Last annotation update)
DE Chymotrypsin/elastase isoform 2 to 5.
OS Ascaris suum (Pig roundworm) (Ascaris lumbricoides).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Ascarididae; Ascaridoidea;
OC Ascarididae; Ascaris.
OX NCBI_TaxID=6253;
RN [1]
RP SEQUENCE.
RX MEDLINE=84255715; PubMed=6564898;
RA Babin D.R., Peanasky R.J., Goos S.M.;
RT "The isoforms of chymotrypsin/elastase from Ascaris
lumbicoides: the primary structure."
RL Arch. Biochem. Biophys. 232:143-161 (1984).
CC -!- FUNCTION: Defend the organism against the host's proteinases.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Contains 1 TIL (Trypsin inhibitory-like) domain.
DR HSSP; P07851; 1EAI.
DR InterPro: IPR002919; TIL_Cysrich.
DR Pfam; PF01826; TIL; 1.
KW Serine protease inhibitor.
FT DOMAIN 4 59
FT DISULFID 4 37 TIL.
FT DISULFID 13 32 BY SIMILARITY.
FT DISULFID 16 28 BY SIMILARITY.
FT DISULFID 20 59 BY SIMILARITY.
FT DISULFID 39 53 BY SIMILARITY.
FT ACT_SITE 30 31 REACTIVE BOND.
FT VARIANT 25 25 K -> N (IN INHIBITOR 2 AND 4).
FT VARIANT 40 40 T -> S (IN INHIBITOR 2 AND 4).
FT VARIANT 64 65 MISSING (IN INHIBITOR 2).
FT VARIANT 65 65 K -> E (IN INHIBITOR 3).
FT VARIANT 65 65 K -> R.
SQ SEQUENCE 65 AA; 7241 MW; B4E51CA166EA4BE3 CRC64;
```

```
Query Match 20.1%; Score 78.5; DB 1; Length 65;
Best Local Similarity 31.7%; Pred. No. 0.04;
Matches 19; Conservative 6; Mismatches 32; Indels 3; Gaps 2;
```

```
QY 9 ROKCPSENFISRCDCGRCQFCFNVVVKPLCKIKICAFGCVCR--LGYLRNKKKVCVPRSKC 66
DB 1 RKPCGKNEVWTECTG-CEKCKQDEKTPCALMCRPFSCCTPGRMWRTHDCKCVPVSEC 59
```

RESULT 4

```
MT_PERFL
ID MT_PERFL STANDARD; PRT; 60 AA.
AC P52725;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Metallothionein (MT).
GN MT.
OS Perca fluviatilis (Perch).
```

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 CC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Percoidae;
 CC Percidae; Percia.
 OX NCBI_TaxID=8168;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RA Kille P., Olsson P.E.;
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE
 CC -!- RESIDUES THAT BIND VARIOUS HEAVY METALS (BY SIMILARITY).
 CC -!- DOMAIN: CLASS I METALLOTHIONEINS CONTAIN 2 METAL-BINDING DOMAINS:
 CC FOUR DIVALENT IONS ARE CHELATED WITHIN CLUSTER A OF THE ALPHA
 CC DOMAIN AND ARE COORDINATED VIA CYSTEINYL THIOLATE BRIDGES TO 11
 CC CYSTEINE LIGANDS. CLUSTER B, THE CORRESPONDING REGION WITHIN THE
 CC BETA DOMAIN, CAN LIGATE THREE DIVALENT IONS TO 9 CYSTEINES.
 CC -!- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 1.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X97272; CRA65927.1; --
 CC HSSP; P28184; 1J19.
 DR InterPro; IPR003019; Metallothion.
 DR InterPro; IPR000006; Metallothion_1.
 DR Pfam; PF00131; metalthio; 1.
 DR PROSITE; PS00203; METALLOTHIONEIN VRT; 1.
 KW Metal-binding; Metal-thiolate cluster.
 FT DOMAIN 1 28 BETA.
 FT DOMAIN 29 60 ALPHA.
 FT METAL 4 4 CLUSTER B.
 FT METAL 6 6 CLUSTER B.
 FT METAL 12 12 CLUSTER B.
 FT METAL 14 14 CLUSTER B.
 FT METAL 18 18 CLUSTER B.
 FT METAL 20 20 CLUSTER B.
 FT METAL 23 23 CLUSTER B.
 FT METAL 25 25 CLUSTER B.
 FT METAL 28 28 CLUSTER B.
 FT METAL 32 32 CLUSTER A.
 FT METAL 33 33 CLUSTER A.
 FT METAL 35 35 CLUSTER A.
 FT METAL 36 36 CLUSTER A.
 FT METAL 40 40 CLUSTER A.
 FT METAL 43 43 CLUSTER A.
 FT METAL 47 47 CLUSTER A.
 FT METAL 49 49 CLUSTER A.
 FT METAL 54 54 CLUSTER A.
 FT METAL 58 58 CLUSTER A.
 FT METAL 59 59 CLUSTER A.
 SQ SEQUENCE 60 AA; 5926 MW; E862A5EC16D7348A CRC64;

 Query Match 18.7%; Score 73; DB 1; Length 60;
 Best Local Similarity 35.0%; Pred. No. 0.15;
 Matches 21; Conservative 3; Mismatches 22; Indels 14; Gaps 3;

 QY 1 GGFGGLGRGKCPSEIFSRCDGRCQRCFQPNVVKPLCIKICAGCYCRLGYNKKKVC 60
 Db 9 GGTCTCGSGCTC-TNCSCTTCKKSCCPGSGCPK-----CAGSGVC-----KGKTC 54

 RESULT 5
 ID -ITR1_ASCSU STANDARD; PRT; 62 AA.
 AC P19398;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Trypsin inhibitor (ATI).
 OS Ascaris suum (Pig roundworm) (Ascaris lumbricoidea).
 CC Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea;
 CC Ascarididae; Ascaris.
 OX NCBI_TaxID=6253;
 RN [1]
 RP SEQUENCE.
 RA Peanasky R.J., Martzen M.R., Homandberg G.A., Cash J.M., Babin D.R.,
 RA Litwiler R.D.;
 RL (In) McInnis A.J. (eds.);
 RL Paradigms for eradicating helminthic parasites, pp.349-366,
 RL Alan R. Liss, New York (1987).
 RN [2]
 RP STRUCTURE BY NMR.
 RX MEDLINE=90212567; PubMed=2322539;
 RA Gronenborn A.M., Nilges M., Peanasky R.J., Clore G.M.;
 RT "Sequential resonance assignment and secondary structure
 RT determination of the Ascaris trypsin inhibitor, a member of a novel
 RT class of proteinase inhibitors.";
 RL Biochemistry 29:183-189(1990).
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- SIMILARITY: Contains 1 TIL (Trypsin inhibitory-like) domain.
 DR PIR; S35098;
 DR PDB; 1ATA; 31-AUG-94.
 DR PDB; 1ATB; 31-AUG-94.
 DR PDB; 1ATD; 31-AUG-94.
 DR PDB; 1ATE; 31-AUG-94.
 DR InterPro; IPR002919; TIL_Cysrich.
 DR Pfam; PF01826; TIL; 1.
 KW Serine protease inhibitor; 3D-structure.
 FT DOMAIN 5 60 TIL.
 FT DISULFID 5 38
 FT DISULFID 15 33
 FT DISULFID 18 29
 FT DISULFID 22 60
 FT DISULFID 40 54
 FT STRAND 11 13
 FT STRAND 16 16
 FT STRAND 34 34
 FT STRAND 37 39
 FT TURN 42 45
 FT STRAND 46 49
 FT TURN 50 52
 FT STRAND 53 56
 FT TURN 58 59
 SQ SEQUENCE 62 AA; 6798 MW; 6438D25D2F76D3E1 CRC64;

 Query Match 18.0%; Score 70.5; DB 1; Length 62;
 Best Local Similarity 29.5%; Pred. No. 0.28;
 Matches 18; Conservative 12; Mismatches 22; Indels 9; Gaps 5;

 QY 11 KCPS-NEIFSRCDGRCQRCFQPNVVKPLCIKICAP---GCVRLGYNKKKVCVPRSK 65
 Db 4 KCTKPNEQWTKCG-CFGTCAQKIVP---CTRECKPFCRCIASAGFVRDAQGNCKIFED 59

 QY 66 C 66
 Db 60 C 60

 RESULT 6
 ID -MTB_ONCMY STANDARD; PRT; 60 AA.
 AC P09862;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Metallothionein B (MT-B).
 GN MTB.
 OS Onchorynchus mykiss (Rainbow trout) (Salmo gairdneri), and
 OS Salvelinus alpinus (Arctic char).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;


```

RT "Molecular cloning of metallothionein cDNA and analysis of
RL metallothionein gene expression in winter flounder tissues.";
CC Can. J. Zool. 67:2520-2527(1989).
CC -1- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE
CC RESIDUES THAT BIND VARIOUS HEAVY METALS (BY SIMILARITY).
CC -1- DOMAIN: CLASS I METALLOTHIONEINS CONTAIN 2 METAL-BINDING DOMAINS:
CC FOUR DIVALENT IONS ARE CHELATED WITHIN CLUSTER A OF THE ALPHA
CC DOMAIN AND ARE COORDINATED VIA CYSTEINYL THIOLATE BRIDGES TO 11
CC CYSTEINE LIGANDS. CLUSTER B, THE CORRESPONDING REGION WITHIN THE
CC BETA DOMAIN, CAN LIGATE THREE DIVALENT IONS TO 9 CYSTEINES.
CC -1- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 1.
CC HSSP; P28184; 1J19.
DR DR InterPro; IPR003019; Metallothion.
DR DR InterPro; IPR000006; Metallothion_1.
DR DR Pfam; PF00131; metalthio; 1.
KW PROSITE; PS00203; METALLOTHIONEIN_VRT; 1.
FT Metal-binding; Metal-thiolate cluster; Copper; Acetylation.
FT MOD RES 1 1 ACETYLATION.
FT DOMAIN 1 28 BETA.
FT DOMAIN 29 60 ALPHA.
FT METAL 4 4 CLUSTER B (BY SIMILARITY).
FT METAL 6 6 CLUSTER B (BY SIMILARITY).
FT METAL 12 12 CLUSTER B (BY SIMILARITY).
FT METAL 14 14 CLUSTER B (BY SIMILARITY).
FT METAL 18 18 CLUSTER B (BY SIMILARITY).
FT METAL 20 20 CLUSTER B (BY SIMILARITY).
FT METAL 23 23 CLUSTER B (BY SIMILARITY).
FT METAL 25 25 CLUSTER B (BY SIMILARITY).
FT METAL 28 28 CLUSTER B (BY SIMILARITY).
FT METAL 32 32 CLUSTER A (BY SIMILARITY).
FT METAL 33 33 CLUSTER A (BY SIMILARITY).
FT METAL 35 35 CLUSTER A (BY SIMILARITY).
FT METAL 36 36 CLUSTER A (BY SIMILARITY).
FT METAL 40 40 CLUSTER A (BY SIMILARITY).
FT METAL 43 43 CLUSTER A (BY SIMILARITY).
FT METAL 47 47 CLUSTER A (BY SIMILARITY).
FT METAL 49 49 CLUSTER A (BY SIMILARITY).
FT METAL 54 54 CLUSTER A (BY SIMILARITY).
FT METAL 58 58 CLUSTER A (BY SIMILARITY).
FT METAL 59 59 CLUSTER A (BY SIMILARITY).
SQ SEQUENCE 60 AA; 6043 MW; 67B046EE5C7470C CRC64;

Query Match 17.4%; Score 68; DB 1; Length 60;
Best Local Similarity 35.2%; Pred.No. 0.5;
Matches 19; Conservative 3; Mismatches 18; Indels 14; Gaps 3;

QY 7 GGRGKCPSEIFSRCDGRQCRFPNVVPKPLCIKICAPGCVCRGLYLRNKKVC 60
DB 15 GGSCTC-KNGSCTTCKNSCCPCPCGCFK-----CASGCV-----KGKTC 54

RESULT 10
MT4 HUMAN
ID MT4 HUMAN STANDARD; PRT; 62 AA.
AC PA7944;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Metallothionein-IV (MT-IV).
GN MT4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94271779; PubMed=8003488;
RA Quafe C.J., Findley S.D., Erickson J.C., Froelick G.J.,
RA Kelly E.J., Zambrowicz B.P., Palmiter R.D.;
RT "Induction of a new metallothionein isoform (MT-IV) occurs during
RL differentiation of stratified squamous epithelia.";
RL Biochemistry 33:7250-7259(1994).
CC -1- FUNCTION: SEEMS TO BIND ZINC AND COPPER. COULD PLAY A SPECIAL ROLE

IN REGULATING ZINC METABOLISM DURING THE DIFFERENTIATION OF
STRATIFIED EPITHELIA.
-1- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 1.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; U07807; AAA20232.1; -.
PIR; A53640; A53640.
HSSP; P18055; IMRB.
DR Genew; HGNC:18705; MT4.
DR MIM; 606206; -.
DR GO; GO:0005737; C:cytoplasm; NAS.
DR GO; GO:0005507; F:copper ion binding activity; NAS.
DR GO; GO:0008270; F:zinc ion binding activity; NAS.
DR InterPro; IPR003019; Metalthion.
DR Pfam; PF00131; metalthio; 1.
DR PRINTS; PR00860; MTVERTERATE.
DR PROSITE; PS00203; METALLOTHIONEIN_VRT; 1.
KW Metal-binding; Metal-thiolate cluster; Zinc; Copper.
FT METAL 6 6 CLUSTER B (BY SIMILARITY).
FT METAL 8 8 CLUSTER B (BY SIMILARITY).
FT METAL 14 14 CLUSTER B (BY SIMILARITY).
FT METAL 16 16 CLUSTER B (BY SIMILARITY).
FT METAL 20 20 CLUSTER B (BY SIMILARITY).
FT METAL 22 22 CLUSTER B (BY SIMILARITY).
FT METAL 25 25 CLUSTER B (BY SIMILARITY).
FT METAL 27 27 CLUSTER B (BY SIMILARITY).
FT METAL 30 30 CLUSTER B (BY SIMILARITY).
FT METAL 34 34 CLUSTER A (BY SIMILARITY).
FT METAL 35 35 CLUSTER A (BY SIMILARITY).
FT METAL 37 37 CLUSTER A (BY SIMILARITY).
FT METAL 38 38 CLUSTER A (BY SIMILARITY).
FT METAL 42 42 CLUSTER A (BY SIMILARITY).
FT METAL 45 45 CLUSTER A (BY SIMILARITY).
FT METAL 49 49 CLUSTER A (BY SIMILARITY).
FT METAL 51 51 CLUSTER A (BY SIMILARITY).
FT METAL 58 58 CLUSTER A (BY SIMILARITY).
FT METAL 60 60 CLUSTER A (BY SIMILARITY).
FT METAL 61 61 CLUSTER A (BY SIMILARITY).
SQ SEQUENCE 62 AA; 6419 MW; 36157CBA17BF28CC CRC64;

Query Match 17.1%; Score 67; DB 1; Length 62;
Best Local Similarity 35.1%; Pred.No. 0.66;
Matches 20; Conservative 8; Mismatches 23; Indels 6; Gaps 4;

QY 7 GGRGKCPSEIFSRCDGR-CORFCPNVVPKPLCIKICAPGCVCRGLYLRNKKVCVP 62
DB 11 GGICWGDNDCKCTTCKTCKRSCPCPCP-PCAK-CARGCICKGG---SDRCSCCP 62

RESULT 11
MT BALMY
ID MT BALMY STANDARD; PRT; 61 AA.
AC O18842;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Metallothionein (MT).
OS Balaena mysticetus (Bowhead whale).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Mysticeti; Balaenidae;
OC Balaena.
OX NCBI_TaxID=27602;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;

```

RA Kaysen J., O'Hara T., Goodwin T., Linnehan R., Hammond T.;
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE
 CC RESIDUES THAT BIND VARIOUS HEAVY METALS (By similarity).
 CC -!- SUBUNIT: Monomer (By similarity).
 CC -!- DOMAIN: CLASS I METALLOTHIONEINS CONTAIN 2 METAL-BINDING DOMAINS:
 CC FOUR DIVALENT IONS ARE CHELATED WITHIN CLUSTER A OF THE ALPHA
 CC DOMAIN AND ARE COORDINATED VIA CYSTEINYL THIOLATE BRIDGES TO 11
 CC CYSTEINE LIGANDS. CLUSTER B, THE CORRESPONDING REGION WITHIN THE
 CC BETA DOMAIN, CAN LIGATE THREE DIVALENT IONS TO 9 CYSTEINES.
 CC -!- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 1.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 CC EMBL; AF022117; AAB72006.1; --
 CC HSSP; P02795; 2MHU.
 CC InterPro; IPR003019; Metallthion.
 CC InterPro; IPR000006; Metallthion_1.
 CC Pfam; PF00131; metalthio; 1.
 CC PRINTS; PR00860; MTVERTEBRATE.
 CC PROSITE; PS00203; METALLOTHIONEIN VRT; 1.
 CC Metal-binding; Metal-thiolate cluster; Acetylation.
 CC DOMAIN 1 29 BETA.
 CC METAL 30 61 ALPHA.
 CC METAL 5 5 CLUSTER B (BY SIMILARITY).
 CC METAL 7 7 CLUSTER B (BY SIMILARITY).
 CC METAL 13 13 CLUSTER B (BY SIMILARITY).
 CC METAL 15 15 CLUSTER B (BY SIMILARITY).
 CC METAL 19 19 CLUSTER B (BY SIMILARITY).
 CC METAL 21 21 CLUSTER B (BY SIMILARITY).
 CC METAL 24 24 CLUSTER B (BY SIMILARITY).
 CC METAL 26 26 CLUSTER B (BY SIMILARITY).
 CC METAL 29 29 CLUSTER B (BY SIMILARITY).
 CC METAL 33 33 CLUSTER A (BY SIMILARITY).
 CC METAL 34 34 CLUSTER A (BY SIMILARITY).
 CC METAL 36 36 CLUSTER A (BY SIMILARITY).
 CC METAL 37 37 CLUSTER A (BY SIMILARITY).
 CC METAL 41 41 CLUSTER A (BY SIMILARITY).
 CC METAL 44 44 CLUSTER A (BY SIMILARITY).
 CC METAL 48 48 CLUSTER A (BY SIMILARITY).
 CC METAL 50 50 CLUSTER A (BY SIMILARITY).
 CC METAL 57 57 CLUSTER A (BY SIMILARITY).
 CC METAL 59 59 CLUSTER A (BY SIMILARITY).
 CC METAL 60 60 CLUSTER A (BY SIMILARITY).
 CC MOD RES 1 1 ACETYLATION (BY SIMILARITY).
 CC SEQUENCE 61 AA; 6025 MW; 4D0A7C5E1D23A4A3 CRC64;
 CC
 CC Query Match 17.0%; Score 66.5; DB 1; Length 61;
 CC Best Local Similarity 38.8%; Pred. No. 0.74;
 CC Matches 19; Conservative 2; Mismatches 21; Indels 7; Gaps 3;
 CC
 CC Qy 1 GGFGLGGRGKPSNEIFSRCDGRQPCPNVVPKPLCIKTCAPGCVCR 49
 CC Db 10 GGSCTCAGSKCKECKCTS-CKKSCCSCP-----PGCTK-CAQGCVCVK 51
 CC
 CC RESULT 12
 CC MTB_SALSA
 CC ID MTB_SALSA STANDARD; PRT; 60 AA.
 CC AC P52720;
 CC DT 01-OCT-1996 (Rel. 34, Created)
 CC DT 01-OCT-1996 (Rel. 34, Last sequence update)
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
 CC DE Metallothionein B (MT-B).
 CC GN MTB.
 CC OS Salmo salar (Atlantic salmon).
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
 OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
 OX NCBF_TaxID=8030;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA TISSUE=Liver;
 RC Kille P., Olsson P.E.;
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE
 CC RESIDUES THAT BIND VARIOUS HEAVY METALS (BY SIMILARITY).
 CC -!- DOMAIN: CLASS I METALLOTHIONEINS CONTAIN 2 METAL-BINDING DOMAINS:
 CC FOUR DIVALENT IONS ARE CHELATED WITHIN CLUSTER A OF THE ALPHA
 CC DOMAIN AND ARE COORDINATED VIA CYSTEINYL THIOLATE BRIDGES TO 11
 CC CYSTEINE LIGANDS. CLUSTER B, THE CORRESPONDING REGION WITHIN THE
 CC BETA DOMAIN, CAN LIGATE THREE DIVALENT IONS TO 9 CYSTEINES.
 CC -!- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 1.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 CC EMBL; X97275; CAA65930.1; --
 CC HSSP; P28184; 1J19.
 CC InterPro; IPR003019; Metallthion.
 CC InterPro; IPR000006; Metallthion_1.
 CC Pfam; PF00131; metalthio; 1.
 CC PRINTS; PR00860; MTVERTEBRATE.
 CC PROSITE; PS00203; METALLOTHIONEIN VRT; 1.
 CC Metal-binding; Metal-thiolate cluster.
 CC DOMAIN 1 28 BETA.
 CC METAL 29 60 ALPHA.
 CC METAL 4 4 CLUSTER B.
 CC METAL 6 6 CLUSTER B.
 CC METAL 12 12 CLUSTER B.
 CC METAL 14 14 CLUSTER B.
 CC METAL 18 18 CLUSTER B.
 CC METAL 20 20 CLUSTER B.
 CC METAL 23 23 CLUSTER B.
 CC METAL 25 25 CLUSTER B.
 CC METAL 28 28 CLUSTER B.
 CC METAL 32 32 CLUSTER A.
 CC METAL 33 33 CLUSTER A.
 CC METAL 35 35 CLUSTER A.
 CC METAL 36 36 CLUSTER A.
 CC METAL 40 40 CLUSTER A.
 CC METAL 43 43 CLUSTER A.
 CC METAL 47 47 CLUSTER A.
 CC METAL 49 49 CLUSTER A.
 CC METAL 54 54 CLUSTER A.
 CC METAL 58 58 CLUSTER A.
 CC METAL 59 59 CLUSTER A.
 CC SEQUENCE 60 AA; 5959 MW; 187A9D0FB5BD967 CRC64;
 CC
 CC Query Match 16.9%; Score 66; DB 1; Length 60;
 CC Best Local Similarity 35.2%; Pred. No. 0.82;
 CC Matches 19; Conservative 3; Mismatches 18; Indels 14; Gaps 3;
 CC
 CC Qy 7 GGRGKPSNEIFSRCDGRQPCPNVVPKPLCIKTCAPGCVCRLLGRLNKKKVC 60
 CC Db 15 GGSCKC-ANCACTSKKSCCPGCGSK-----CASGCV-----KGKTC 54
 CC
 CC RESULT 13
 CC MT_ZOAVI
 CC ID MT_ZOAVI STANDARD; PRT; 60 AA.
 CC AC P52728;
 CC DT 01-OCT-1996 (Rel. 34, Created)
 CC DT 01-OCT-1996 (Rel. 34, Last sequence update)
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Metallothionein (MT).

GN MT.

OS Zoarces viviparus (Belpout).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Zoarcoidae;

OC Zoarcidae; Zoarces.

OX NCBI_TaxID=48416;

RN [1]_TaxID=48416;

RP SEQUENCE FROM N.A.

RC TISSUE=Liver;

RA Kille P., Olsson P.B.;

RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE

CC RESIDUES THAT BIND VARIOUS HEAVY METALS (BY SIMILARITY).

CC -!- DOMAIN: CLASS I METALLOTHIONEINS CONTAIN 2 METAL-BINDING DOMAINS:

CC FOUR DIVALENT IONS ARE CHLORATED WITHIN CLUSTER A OF THE ALPHA

CC DOMAIN AND ARE COORDINATED VIA CYSTEINYL THIOALATE BRIDGES TO 11

CC CYSTEINE LIGANDS. CLUSTER B, THE CORRESPONDING REGION WITHIN THE

CC BETA DOMAIN, CAN LIGATE THREE DIVALENT IONS TO 9 CYSTEINES.

CC -!- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 1.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; X97270; CAA65925.1; -.

DR HSP; P28184; 1J19.

DR InterPro; IPR003019; Metallothion.

DR InterPro; IPR000006; Metallothion_1.

DR Pfam; PF00131; metalthio; 1.

DR PROSITE; PS00203; METALLOTHIONEIN VRT; 1.

KW Metal-binding; Metal-thiolate cluster.

FT DOMAIN 1 28 BETA.

FT METAL 29 60 ALPHA.

FT METAL 4 4 CLUSTER B.

FT METAL 6 6 CLUSTER B.

FT METAL 12 12 CLUSTER B.

FT METAL 14 14 CLUSTER B.

FT METAL 18 18 CLUSTER B.

FT METAL 20 20 CLUSTER B.

FT METAL 23 23 CLUSTER B.

FT METAL 25 25 CLUSTER B.

FT METAL 28 28 CLUSTER B.

FT METAL 32 32 CLUSTER A.

FT METAL 33 33 CLUSTER A.

FT METAL 35 35 CLUSTER A.

FT METAL 36 36 CLUSTER A.

FT METAL 40 40 CLUSTER A.

FT METAL 43 43 CLUSTER A.

FT METAL 47 47 CLUSTER A.

FT METAL 49 49 CLUSTER A.

FT METAL 54 54 CLUSTER A.

FT METAL 58 58 CLUSTER A.

FT METAL 59 59 CLUSTER A.

SQ SEQUENCE 60 AA; 6047 MW; E878BD4FE2BC471A CRC64;

Query Match 16.6%; Score 65; DB 1; Length 60;

Best Local Similarity 37.0%; Pred. No. 1.1;

Matches 20; Conservative 3; Mismatches 17; Indels 14; Gaps 4;

QY 7 GGRGKCPSEIFSRCDGRQRCFQPNVVPKLCIKICAPGCVRLGYNKKKVC 60

DB 15 GGSCKG-TNCSCTTKCKSCPCPSG-----CTK-CASGCV-----KGKTC 54

RESULT 14

MT4_MOUSE STANDARD; PRT; 62 AA.

ID MT4_MOUSE

AC P47945;

DT 01-FEB-1996 (Rel. 33, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Metallothionein-IV (MT-IV).

GN MT4.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]_TaxID=10090;

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6;

RX MEDLINE=94371779; PubMed=8003488;

RA Quafe C.J., Findley S.D., Erickson J.C., Froelick G.J.,

RA Kelly E.J., Zambrowicz B.P., Palmer R.D.;

RT "Induction of a new metallothionein isoform (MT-IV) occurs during

RT differentiation of stratified squamous epithelia.";

RL Biochemistry 33:7250-7259(1994).

CC -!- FUNCTION: SEEMS TO BIND ZINC AND COPPER. COULD PLAY A SPECIAL ROLE

CC IN REGULATING ZINC METABOLISM DURING THE DIFFERENTIATION OF

CC STRATIFIED EPITHELIA.

CC -!- TISSUE SPECIFICITY: EXPRESSED EXCLUSIVELY IN STRATIFIED SQUAMOUS

CC EPITHELIA ASSOCIATED WITH ORAL EPITHELIA, OESOPHAGUS, UPPER

CC STOMACH, TAIL, FOOTPADS, AND NEONATAL SKIN.

CC -!- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 1.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; U07808; AAA20233.1; -.

DR PIR; B53640; B53640.

DR HSP; P18055; 1MRB.

DR MGD; MGI:93692; Mt4.

DR InterPro; IPR003019; Metallothion.

DR InterPro; IPR000006; Metallothion_1.

DR Pfam; PF00131; metalthio; 1.

DR PROSITE; PS00203; METALLOTHIONEIN VRT; 1.

KW Metal-binding; Metal-thiolate cluster; zinc; Copper.

FT METAL 6 6 CLUSTER B (BY SIMILARITY).

FT METAL 8 8 CLUSTER B (BY SIMILARITY).

FT METAL 14 14 CLUSTER B (BY SIMILARITY).

FT METAL 16 16 CLUSTER B (BY SIMILARITY).

FT METAL 20 20 CLUSTER B (BY SIMILARITY).

FT METAL 22 22 CLUSTER B (BY SIMILARITY).

FT METAL 25 25 CLUSTER B (BY SIMILARITY).

FT METAL 27 27 CLUSTER B (BY SIMILARITY).

FT METAL 30 30 CLUSTER B (BY SIMILARITY).

FT METAL 34 34 CLUSTER A (BY SIMILARITY).

FT METAL 35 35 CLUSTER A (BY SIMILARITY).

FT METAL 37 37 CLUSTER A (BY SIMILARITY).

FT METAL 38 38 CLUSTER A (BY SIMILARITY).

FT METAL 42 42 CLUSTER A (BY SIMILARITY).

FT METAL 45 45 CLUSTER A (BY SIMILARITY).

FT METAL 49 49 CLUSTER A (BY SIMILARITY).

FT METAL 51 51 CLUSTER A (BY SIMILARITY).

FT METAL 58 58 CLUSTER A (BY SIMILARITY).

FT METAL 60 60 CLUSTER A (BY SIMILARITY).

FT METAL 61 61 CLUSTER A (BY SIMILARITY).

SQ SEQUENCE 62 AA; 6276 MW; 8F78DA6D0744333A CRC64;

Query Match 16.6%; Score 65; DB 1; Length 62;

Best Local Similarity 33.9%; Pred. No. 1.1;

Matches 21; Conservative 4; Mismatches 27; Indels 10; Gaps 4;

QY 1 GGFGLGGRKCPSEIFSRCDGRQRCFQPNVVPKLCIKICAPGCVRLGYNKKKVC 60

DB 11 GGICIGDNCKCTCTCSC-KTCRKSQCCP-----PGCAK-CARGICKGG---SDKSC 60

Qy 61 VP 62
Db 61 CP 62

RESULT 15

MT_ABBREV MT_ABBREV STANDARD; PRT; 60 AA.
AC O42152;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Metallothionein (MT) (AnMT).
GN MT-A.
OS Ambystoma mexicanum (Axolotl).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Caudata; Salamandroidea; Ambystomatidae;
OC Ambystoma.
OX NCBI_TaxID=8296;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=98127743; PubMed=9468225;
RA Saint-Jacques E., Guay J., Wirtanen L., Huard V., Stewart G.,
RA Seguin C.;
RT "Cloning of a complementary DNA encoding an Ambystoma mexicanum
RT metallothionein, AnMT, and expression of the gene during early
RT development.";
RL DNA Cell Biol. 17:83-91(1998).
CC -!- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE
CC RESIDUES THAT BIND VARIOUS HEAVY METALS (By similarity).
CC -!- DOMAIN: CLASS I METALLOTHIONEINS CONTAIN 2 METAL-BINDING DOMAINS:
CC FOUR DIVALENT IONS ARE CHELATED WITHIN CLUSTER A OF THE ALPHA
CC DOMAIN AND ARE COORDINATED VIA CYSTEINYL THIOLATE BRIDGES TO 11
CC CYSTEINE LIGANDS. CLUSTER B, THE CORRESPONDING REGION WITHIN THE
CC BETA DOMAIN, CAN LIGATE THREE DIVALENT IONS TO 9 CYSTEINES.
CC -!- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF008583; AAB71835.1; -
CC HSSP; P02795; 1MHU.
CC InterPro; IPR002400; GF cystknot.
CC InterPro; IPR003019; Metalthion.
CC InterPro; IPR000006; Metalthion_1.
CC Pfam; PF00131; metalthio; 1.
CC PRINTS; PR00438; GFCYSKNOT.
CC PRINTS; PR00860; MTVERTEBRATE.
CC PROSITE; PS00203; METALLOTHIONEIN VRT; 1.
KW Metal-binding; Metal-thiolate cluster.
FT DOMAIN 1 28 BETA.
FT DOMAIN 29 60 ALPHA.
FT METAL 3 3 CLUSTER B.
FT METAL 5 5 CLUSTER B.
FT METAL 11 11 CLUSTER B.
FT METAL 13 13 CLUSTER B.
FT METAL 17 17 CLUSTER B.
FT METAL 19 19 CLUSTER B.
FT METAL 22 22 CLUSTER B.
FT METAL 24 24 CLUSTER B.
FT METAL 27 27 CLUSTER B.
FT METAL 31 31 CLUSTER A.
FT METAL 32 32 CLUSTER A.
FT METAL 34 34 CLUSTER A.
FT METAL 35 35 CLUSTER A.
FT METAL 39 39 CLUSTER A.

FT METAL 42 42 CLUSTER A.
FT METAL 46 46 CLUSTER A.
FT METAL 48 48 CLUSTER A.
FT METAL 56 56 CLUSTER A.
FT METAL 58 58 CLUSTER A.
FT METAL 59 59 CLUSTER A.
SQ SEQUENCE 60 AA; 5979 MW; AB05F49153953279 CRC64;
Query Match 16.2%; Score 63.5; DB 1; Length 60;
Best Local Similarity 33.3%; Pred. No. 1.5;
Matches 17; Conservative 4; Mismatches 23; Indels 7; Gaps 2;
Qy 1 GGFGLGGGKCPSPNEIFSRCDGRQRCFNPVVPKPLKIKICAPCTVCRLG 51
Db 8 GGSCSCAGSCCKCCKTS-CKKSCCCCPSECEK-----CGQGCVCCKGG 51

Search completed: November 17, 2003, 08:16:04
Job time : 11 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2003, 08:13:38 ; Search time 34 Seconds

(without alignments)
508.515 Million cell updates/sec

Title: US-09-506-978-1

Perfect score: 391

Sequence: 1 GGGGLGGGKCPSEIFSR.....CRGLYLRNKKKVCVPRSKCG 67

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 66991

Minimum DB seq length: 0

Maximum DB seq length: 67

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	88.5	22.6	62	5 077419	O77419 ascaris suu
2	63.5	16.2	61	4 Q8WVB5	Q8WVB5 homo sapien
3	63	16.1	51	13 Q9PS78	Q9PS78 gallus gall
4	62.5	16.0	40	13 Q98TP9	Q98TP9 platichthys
5	62.5	16.0	44	11 Q99KF9	Q99KF9 mus musculus
6	60	15.3	61	4 Q8TDN3	Q8TDN3 homo sapien
7	60	15.3	61	4 Q8T339	Q8T339 homo sapien
8	59.5	15.2	63	4 Q9BSK9	Q9BSK9 homo sapien
9	59	15.1	48	13 Q98TC0	Q98TC0 seriola qui
10	59	15.1	61	6 Q8MI14	Q8MI14 bos taurus
11	58.5	15.0	61	4 Q8TE66	Q8TE66 homo sapien
12	58	14.8	60	13 Q8UYV1	Q8UYV1 lithognathu
13	57.5	14.7	49	13 Q98S15	Q98S15 leuciscus c
14	57.5	14.7	60	13 Q8AWG2	Q8AWG2 carassius c
15	57	14.6	57	13 Q8JHC5	Q8JHC5 gobius nige
16	55.5	14.2	60	13 Q8AWG1	Q8AWG1 carassius c

17	55	14.1	55	17	Q8PRS9	Q8PRS9 methanosarc
18	54.5	13.9	48	4	Q12928	Q12928 homo sapien
19	54.5	13.9	61	4	Q8TDC4	Q8TDC4 homo sapien
20	53	13.6	39	13	Q9PVG7	Q9PVG7 xiphophorus
21	53	13.6	59	16	Q8XPB4	Q8XPB4 clostridium
22	53	13.6	66	4	Q8TCS1	Q8TCS1 homo sapien
23	51.5	13.2	59	5	Q9N9H1	Q9N9H1 ruditaes d
24	51.5	13.2	62	10	Q65918	Q65918 picea maria
25	51	13.0	45	4	Q8NEA2	Q8NEA2 homo sapien
26	51	13.0	48	16	Q8YUR2	Q8YUR2 anabaena sp
27	51	12.9	65	17	Q8TNN8	Q8TNN8 methanosarc
28	50.5	12.9	58	5	Q95P38	Q95P38 homarus ame
29	50	12.8	58	5	Q95U91	Q95U91 scylla serr
30	50	12.8	67	16	Q8ZKI6	Q8ZKI6 salmoneilla
31	49.5	12.7	55	17	Q97BH4	Q97BH4 thermoplas
32	49.5	12.7	58	5	Q9U623	Q9U623 pacifastacu
33	49	12.5	43	4	Q8NES8	Q8NES8 homo sapien
34	49	12.5	58	5	Q95U93	Q95U93 portunus pe
35	49	12.5	58	5	Q95U92	Q95U92 eriocheir s
36	49	12.5	59	5	Q9TWP8	Q9TWP8 anemonia su
37	48.5	12.4	29	10	Q988D2	Q988D2 cucumis mel
38	48.5	12.4	31	6	Q77625	Q77625 bos taurus
39	48.5	12.4	41	4	Q9UDP7	Q9UDP7 homo sapien
40	48.5	12.4	55	17	Q9HJ78	Q9HJ78 thermoplas
41	47.5	12.1	37	11	Q9QX87	Q9QX87 rattus norv
42	47.5	12.1	57	5	Q9N9H2	Q9N9H2 venerupis (
43	47.5	12.1	61	4	Q9BXG3	Q9BXG3 homo sapien
44	47.5	12.1	61	6	Q18780	Q18780 ovis aries
45	47	12.0	41	13	Q9PVG6	Q9PVG6 xiphophorus

ALIGNMENTS

RESULT 1

O77419 PRELIMINARY; PRT; 62 AA.
AC O77419;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DB Chymotrypsin/elastase inhibitor-1 (Fragment).
GN ASC/E-1.
OS Ascaris suum (Pig roundworm) (Ascaris lumbricoidea).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea;
OC Ascarididae; Ascaris.
OX NCBI_TaxID=6253;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98297373; PubMed=9635450;
RA Lu C.C., Nguyen T., Morris S., Hill D., Sakanari J.A.;
RT "Anisakis simplex: mutational bursts in the reactive site centers of
serine protease inhibitors from an ascarid nematode."
RL Exp. Parasitol. 89:257-261(1998).
DR EMBL; U94499; AAC61300.1; -
DR HSSP; P07851; IEAI
DR InterPro; IPR002919; TIL_Cysrich.
DR Pfam; PF01826; TIL; 1.
FT NON TER 1
FT NON TER 62
SQ SEQUENCE 62 AA; 6794 MW; 154CE25A375E0B2C CRC64;

Query Match 22.6%; Score 88.5; DB 5; Length 62;
Best Local Similarity 33.3%; Pred. No. 0.00076;
Matches 21; Conservative 9; Mismatches 26; Indels 7; Gaps 4;

Qy 8 GRGKCPNEIFSCDRCQRCF--PNVVPKPLIKIKCAPCVCVCR--LGYLNRKKKVCVPR 63
Db 1 GQRGCGNEVWTECTG--CEMKCGDPENTPCPLMCR--RPSCEGSGRGMRTNDGKICPA 57
Qy 64 SKC 66
Db 58 SQC 60

RESULT 2

Q8WB55 ID Q8WB55 PRELIMINARY; PRT; 61 AA.
AC Q8WB55;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to metallothionein 1L (MT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE
CC RESIDUES THAT BIND VARIOUS HEAVY METALS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 1.
DR EMBL; BC018190; AAH18190.1; -
DR InterPro; IPR003019; Metallothion_1.
DR Pfam; PF00131; metalthio; 1.
DR PRINTS; PR00860; MTVERTEBRATE.
DR PROSITE; PS00203; METALLOTHIONEIN_VRT; 1.
KW Metal-binding; Metal-thiolate cluster.
SQ SEQUENCE 61 AA; 6054 MW; 3C4827F942E986B5 CRC64;

Query Match 16.2%; Score 63.5; DB 4; Length 61;
Best Local Similarity 34.8%; Pred. No. 1.1; Mismatches 6; Gaps 4;
Matches 16; Conservative 6; Indels 5; Gaps 4;
QY 4 GGLGGKCPSPNEIFSRCDGRQCFNVPVVKPLCIKICAPGCVCR 49
Db 11 GSCACAGSCKNE--CKTS-CKKSCCCPVG-CAK-CAQGCLCK 51

RESULT 3

Q9PS78 ID Q9PS78 PRELIMINARY; PRT; 51 AA.
AC Q9PS78;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Insulin-like growth factor binding protein-5, IGFBP-5 (Fragment).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96414636; PubMed=8817657;
RA Allander S.V., Ehrenberg E., Luthman H., Powell D.R.;
RT "Conservation of IGFBP structure during evolution: cloning of chicken
RT insulin-like growth factor binding protein-5.";
RL Prog. Growth Factor Res. 6:159-165(1995).
DR InterPro; IPR000867; Insl_gro_fac_pr.
DR Pfam; PF00219; IGFBP; 1.
DR SMART; SM00121; IB; 1.
DR PROSITE; PS00222; IGF_BINDING; 1.
FT NON_TER 1
FT NON_TER 51
SQ SEQUENCE 51 AA; 5127 MW; 662EA453F4CCB267 CRC64;
Query Match 16.1%; Score 63; DB 13; Length 51;
Best Local Similarity 45.2%; Pred. No. 1.1; Mismatches 14; Conservative 2; Indels 4; Gaps 2;
QY 21 CDGRQCFNVPVVKPL-CLKICAPGCVCL 50

Db 10 CDGRKALSCLP---PPPLGCELVKPEPGCGCL 37

RESULT 4

Q98TP9 ID Q98TP9 PRELIMINARY; PRT; 40 AA.
AC Q98TP9;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Metallothionein (MT) (Fragment).
GN MT.
OS Platicthys flesus (European flounder).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
OC Pleuronectoidae; Pleuronectidae; Platicthys.
OX NCBI_TaxID=8260;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RA Williams T.D., Chipman J.K.;
RT "A DNA array to monitor the effects of environmental pollution on
RT European flounder (Platicthys flesus).";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE
CC RESIDUES THAT BIND VARIOUS HEAVY METALS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 1.
DR EMBL; AJ291833; CAC28138.1; -
DR HSP; P02795; IMHU.
DR InterPro; IPR003019; Metallothion.
DR InterPro; IPR000006; Metallothion_1.
DR Pfam; PF00131; metalthio; 1.
DR PROSITE; PS00203; METALLOTHIONEIN_VRT; 1.
KW Metal-binding; Metal-thiolate cluster.
FT NON_TER 1
FT NON_TER 40
SQ SEQUENCE 40 AA; 4010 MW; B5CEB05E4200855B CRC64;

Query Match 16.0%; Score 62.5; DB 13; Length 40;
Best Local Similarity 37.2%; Pred. No. 1; Mismatches 16; Conservative 3; Indels 7; Gaps 2;
Matches 16; Conservative 3; Mismatches 17; Indels 7; Gaps 2;

QY 7 GGRGKCPSPNEIFSRCDGRQCFNVPVVKPLCIKICAPGCVCR 49
Db 5 GGSCTC-KNCSCTTNKSCCPGFCPK-----CAGCGVCK 40

RESULT 5

Q99KF9 ID Q99KF9 PRELIMINARY; PRT; 44 AA.
AC Q99KF9;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical 4.9 kDa protein.
GN 3110001K13RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC004681; AAH04681.1; -
DR MGD; MGI:1919666; 3110001K13RIK.
KW Hypothetical protein.
SQ SEQUENCE 44 AA; 4900 MW; C49523A82D448591 CRC64;
Query Match 16.0%; Score 62.5; DB 11; Length 44;
Best Local Similarity 33.3%; Pred. No. 1.1;

Matches 14; Conservative 5; Mismatches 14; Indels 9; Gaps 2;

Qy	:	25	CQRF	PVVPKPLCIKICAPGCVCLGYLRNKKVCVPRSKC	66
	:			:	
Db	:	10	CLCVCSV-----CVSVCSVCVCLCVYMR----	AHMPMSKC	42

RESULT 6
Q8TDN3
ID Q8TDN3 PRELIMINARY; PRT; 61 AA.

AC	01-JUN-2002	(TReMBLrel. 21, Created)
DT	01-JUN-2002	(TReMBLrel. 21, Last sequence update)
DT	01-MAR-2003	(TReMBLrel. 23, Last annotation update)

DE Metallothionein in (M1).
GN MTM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eu

```

OC Mammalia; Eutneria; Filimates; Catafrimnia; Homimilidae;
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N A

```

Wang J., Yu L., Zhao S.
"Cloning of a novel member of the MT gene family-MT1M.
Submitted (FEB-2001) to the EMBL/GenBank/DBJ database
FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF

CC
CC
CC
CC
DR
DR
DR

-!- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY
EMBL; AF348671; AAL83902.1; -.
InterPro: IPR002400. CF structnet.

DR InterPro; IPR003019; Metallthion.
DR InterPro; IPR000006; Metallthion₁.
DR Pfam; PF00131; metalthio; 1.
DR PROSITE; PS00428; Cysprolylase

DR PROSITE; PS00203; METALLOTHIONEIN_VRT; 1.
KW Metal-binding; Metal-thiolate cluster.
SQ SEQUENCE 61 AA; 6110 MW; 754753ED088411CE CRC64;

Query Match 15.3%; Score 60; DB 4; Length
Best Local Similarity 39.1%; Pred. No. 3.2;
Matches 18; Conservative 5; Mismatches 17; Inde

QY | | | | | : | : | : | : | : | :
10 GKCPSEI FSRCDGR CQRFC PNVPV PKPLTKICAPGCVCRLGYLR

Db | | | : | : | : | : | : | : | :
17 GSCTCKE --CKTCS-CKKSCCSCPVG-CAK-CAHGCVCVK-GTLE

RESULT 7
Q8N339
ID Q8N339 PRELIMINARY; PRT; 61 AA.

DT	(TREM)rel.	22,	Created)
DT	01-OCT-2002	(TREM)rel.	22, Created)
DT	01-OCT-2002	(TREM)rel.	22, Last sequence update)
DT	01-MAR-2003	(TREM)rel.	23, Last annotation update)

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Eu-
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; H-

CA RN [1] NCBI_taxID=9606;
RP SEQUENCE FROM N.A.
RC TISSUE=Colon. and Kidney;

RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ database
DR EMBL; BC028280; AA28280.1; -.
DP InterPro: IPR002400; C5 cysteine+

```

DR InterPro; IPR003019; Metallthion.
DR InterPro; IPR000006; Metallthion_1.
DR Pfam; PF00131; metalthio; 1.
DR PRINTS; PR00430; CTRCROX20M

```

DR PRINTS; PR00860; MTVERTEBRATE.
DR PROSITE; PS00203; METALLOTHIONEIN_VRT; 1.
KW Hypothetical protein.

SQ SEQUENCE 61 AA; 6137 MW; 755E7B4F9B8411CE CRC64;

Query Match 15.3%; Score 60; DB 4; Length 61;
Best Local Similarity 39.1%; Pred. No. 3.2;
Matches 18; Conservative 5; Mismatches 17; Indels

Qy 10 GKPSNEIFSRCDGRQRFCEWVVKPLCIKICAPGCVCRLGYLRN 55
Db 17 GSKCKKE--CKCTS-CKKSCTCCCPVG-CAK-CAHGCVVK-GTLEN 56

RESULT 8
Q9BSK9
ID Q9BSK9 PRELIMINARY: PRT: 63 AA.

AC	Q958K3	DT	01-JUN-2001	(TrEMBLrel. 17, Created)
DT	01-JUN-2001	(TrEMBLrel. 17, Last sequence update)		
DT	01-OCT-2002	(TrEMBLrel. 22, Last annotation update)		

DE Hypotnetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Eutele
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
PC TISSUE=Skin.

RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC004969; AAH04969.1; -.
Unpublished

SQ	SEQUENCE	63 AA; 8DP7AB2529D8628C CRC64;
	Query Match	15.2%; Score 59.5; DB 4; Length 63
	Identical	20.4%
	Similarity	20.4%

Qy	Matches	15; Conservative	4; Mismatches	19; Indels
2	GFGGLGGRGKCP	NEIFSR	CDGRFC	PNVVPKPLCIK-----ICAP

Db 15 GFSGVLGGRCQG-----GSCESLGPPPPASSLLCSKPASGPMVCLE

ID	Q98TC0	PRELIMINARY;	PRT;	48 AA.
AC	Q98TC0;			
DT	01-JUN-2001	(T=EMBLrel. 17, Created)		

DE Metallothionein (MT) (Fragment).
GN MT.

[illegible]

CC Carangidae, Serridae.
OX NCBI_TaxID=8161;
RN [1]
RP SEQUENCE FROM N. A.

RT "Molecular cloning of metallothionein cDNA from yellow tai
RT quinquerediata) for monitoring of environmental pollution.
PI. Submitted (Apr-2001) to the EMBL/GenBank/DBI databases
RA Futami K., Maeta M.;

CC RESIDUES THAT BIND VARIOUS HEAVY METALS (BY SIMILARITY
CC -I- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTE
CC -I- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY
CC EMPID: AB050242, BAB41015, 1-

DR HSSP; P02795; 1MHU.
DR InterPro; IPR003019; Metallthion.
DR InterPro; IPR000006; Metallthion_1.
DR InterPro; IPR000006; Metallthion_1.
DR InterPro; IPR000006; Metallthion_1.

DR PROSITE; PS00203; METALLOTHIONEIN_VRT; 1.
KW Metal-binding; Metal-thiolate cluster.
FT NON_TER 1 1


```
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Hayes R.A., Gensberg K., Clark A., Williams T.D.;
RT "Chub metallothionein as a bioindicator of heavy metal pollution.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE
CC RESIDUES THAT BIND VARIOUS HEAVY METALS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 1.
CC EMBL; AY029370; AAU31301.1; -.
DR HSP; P02795; IMHU.
DR InterPro; IPR003019; Metallothion.
DR InterPro; IPR000006; Metallothion_1.
DR Pfam; PF00131; metalthio; 1.
DR PROSITE; PS00203; METALLOTHIONEIN_VRT; 1.
KW Metal-binding; Metal-thiolate cluster.
FT NON_TER 1
FT NON_TER 49
SQ SEQUENCE 49 AA; 4803 MW; AFDCT7534D883CBA CRC64;

Query Match 14.7%; Score 57.5; DB 13; Length 49;
Best Local Similarity 34.9%; Pred. No. 5.4;
Matches 15; Conservative 4; Mismatches 17; Indels 7; Gaps 2;

QY 7 GGRGKCPSEIFSRCDGRQCFPCNVVFKPLCIKICAPGCVCR 49
Db | | | | | | | | | | | | | | | | | | | |
9 GATCKC-TNCQCTTKKSCCTCCPSGCSK-----CASGGCVCK 44

RESULT 14
Q8AWG2 PRELIMINARY; PRT; 60 AA.
AC Q8AWG2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Metallothionein.
OS Carassius cuvieri.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Carassius.
CX NCBI_TaxID=52617;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2016892; PubMed=10706375;
RA Ren H.W., Itoh N., Kanekiyo M., Tominaga S., Kohroki J., Hwang G.S.,
RA Nakanishi T., Muto N., Tanaka K.;
RT "Two metallothioneins in the fresh-water fish, crucian carp (Carassius
RT cuvieri): cDNA cloning and assignment of their expression isoforms.";
RL Biol. Pharm. Bull. 23:145-148(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Ren H.W., Itoh N., Kanekiyo M., Tominaga S., Kohroki J., Hwang G.S.,
RA Nakanishi T., Muto N., Tanaka K.;
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY165047; AAN85819.1; -.
SQ SEQUENCE 60 AA; 5958 MW; 543F9410864BB701 CRC64;

Query Match 14.7%; Score 57.5; DB 13; Length 60;
Best Local Similarity 34.9%; Pred. No. 6.5;
Matches 15; Conservative 4; Mismatches 17; Indels 7; Gaps 2;

QY 7 GGRGKCPSEIFSRCDGRQCFPCNVVFKPLCIKICAPGCVCR 49
Db | | | | | | | | | | | | | | | | | | | |
15 GATCKC-TNCQCTTKKSCCSCPSGCSK-----CASGGCVCK 50

RESULT 15
Q8JHCS PRELIMINARY; PRT; 57 AA.
ID Q8JHCS
AC Q8JHCS;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
```

```
DE Metallothionein (Fragment).
OS Gobius niger (Black goby).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Gobioidae;
OC Gobiidae; Gobius.
CX NCBI_TaxID=85417;
RN [1]
RP SEQUENCE FROM N.A.
RA Maradonna F., Cardinali M., Carnevali O.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF520609; AAM74951.1; -.
DR InterPro; IPR003019; Metallothion.
DR InterPro; IPR000006; Metallothion_1.
DR Pfam; PF00131; metalthio; 1.
DR PRINTS; PR00860; MTVERTEBRATE.
DR PROSITE; PS00203; METALLOTHIONEIN_VRT; 1.
FT NON_TER 1
FT NON_TER 57
SQ SEQUENCE 57 AA; 5603 MW; 67D04FEC1456159C CRC64;

Query Match 14.6%; Score 57; DB 13; Length 57;
Best Local Similarity 33.3%; Pred. No. 7.1;
Matches 18; Conservative 2; Mismatches 20; Indels 14; Gaps 3;

QY 7 GGRGKCPSEIFSRCDGRQCFPCNVVFKPLCIKICAPGCVCR 60
Db | | | | | | | | | | | | | | | | | | | |
14 GGSCTC-TNCSTCKKSCCPCGCSK-----CASGGCVCK 53

Search completed: November 17, 2003, 08:16:52
Job time : 35 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2003, 08:09:58 ; Search time 41 seconds
(without alignments)
259.382 Million cell updates/sec

Title: US-09-506-978-1
Perfect score: 391
Sequence: 1 GFGGLGGRKCPSPNEIFSR.....CRLLGLYLRNKKVCVPRSKCG 67

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 631746

Minimum DB seq length: 0
Maximum DB seq length: 67

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

```

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	391	100.0	67	23 ABB08330	Bee venom protein
2	82	21.0	61	20 AAY30434	Mature nematode ex
3	82	21.0	61	21 AAB15319	N. americana nemat
4	68	17.4	58	20 AAY30433	Mature nematode ex
5	63.5	16.2	59	21 AAY57812	Trout metallothion
6	63.5	16.2	59	23 ABP32832	Human ORF1805 prot
7	62.5	16.0	62	23 ABB77804	Consensus sequenc
8	62.5	16.0	63	23 AAO21321	Arabidopsis thalia
9	62	15.9	43	22 AAB60286	Human factor IX (h

10	62	15.9	61	21 AAY57822	Rabbit liver metal
11	61.5	15.7	67	22 AAB61455	Metallothionein do
12	61	15.6	47	20 AAY04648	Factor IXa catalyt
13	61	15.6	56	21 AAB44778	Human secreted pro
14	61	15.6	66	22 AAU45263	Propionibacterium
15	60	15.3	60	23 ABP01364	Human ORFX protein
16	60	15.3	61	24 ABP97389	Human metallothion
17	59.5	15.2	61	22 AAU28133	Novel human secret
18	59	15.1	61	19 AAW61601	Human metallothion
19	58.5	15.0	39	20 AAY02082	KTX peptide used t
20	58.5	15.0	61	22 AAM40263	Human polypeptide
21	58.5	15.0	61	23 ABP65139	Hypoxia-regulated
22	58	14.8	56	22 AAU48798	Propionibacterium
23	57.5	14.7	61	23 ABP65140	Hypoxia-regulated
24	57.5	14.7	67	22 AAM79414	Human protein SEQ
25	57	14.6	10	21 AAY69218	N-terminal sequenc
26	57	14.6	40	5 AAP40692	Sequence encoded b
27	57	14.6	40	5 AAP40220	Partial sequence o
28	56.5	14.5	55	22 AAM39148	Human polypeptide
29	56.5	14.5	60	14 AAR40209	Sequence of human
30	56.5	14.5	60	21 AAY82332	Human metallothion
31	56.5	14.5	61	21 AAB12587	Human metallothion
32	56.5	14.5	61	21 AAY82331	Human metallothion
33	56.5	14.5	61	22 AAM78430	Human protein SEQ
34	56.5	14.5	61	23 ABP65151	Hypoxia-regulated
35	56.5	14.5	61	23 ABB09810	Amino acid sequenc
36	56	14.3	66	23 ABP02424	Human ORFX protein
37	55.5	14.2	46	21 AAB12508	Beta-cellulin mute
38	55.5	14.2	47	21 AAB12507	Beta-cellulin mute
39	55.5	14.2	47	21 AAB12516	Beta-cellulin mute
40	55.5	14.2	48	21 AAB12515	Beta-cellulin mute
41	55.5	14.2	48	21 AAB12517	Beta-cellulin mute
42	55.5	14.2	48	21 AAB12518	Beta-cellulin mute
43	55.5	14.2	53	21 AAB12521	Beta-cellulin mute
44	55.5	14.2	60	21 AAM49082	Metallothionein fr
45	55.5	14.2	61	23 AAE28599	Metallothionein 1a

ALIGNMENTS

RESULT 1

ABB08330
ID ABB08330 standard; protein; 67 AA.

XX ABB08330;

XX 18-JUN-2002 (first entry)

XX Bee venom protein Api m 6.01.

XX Bee venom, isoform; immunosuppressant; vaccine; Api m 6; immune response;
XX bee venom hypersensitivity; antibody; protein purification; Api m 6.01;
XX immunotherapy; allergen.

OS Apis sp.

XX WO200188085-A2.

XX 22-NOV-2001.

XX 16-FEB-2001; 2001WO-IB01736.

XX 18-FEB-2000; 2000US-0506978.

XX (ECOL-) ECOLE POLYTECHNIQUE FEDERALE LAUSANNE.

XX Spertini F;

XX WPI; 2002-082988/11.

XX New bee venom polypeptides, useful for modulating immune responses e.g.
XX in individual hypersensitive to the venom and for identifying

PT individual at risk for bee venom hypersensitivity -
XX Example 2; Page 26; 32pp; English.
XX The present sequence is that of one the four isoforms of Api m 6,
CC designated Api m 6.01. The sequence represents the central amino acid
CC sequence shared by all four isoforms (see ABB08331, ABB08332 and
CC ABB08333). The specification describes a substantially pure polypeptide,
CC Api m 6, derived from bee venom and found in four isoforms. The proteins
CC of the invention have immunosuppressant activity and may form the basis
CC as a vaccine. Api m 6 is useful for modulating an immune response, i.e.
CC as an allergen for immunotherapy. The protein is useful for identifying
CC an individual at risk for bee venom hypersensitivity. The method
CC comprises administering Api m 6 to the individual and measuring an immune
CC response raised, where a detectable immune response indicates that the
CC individual is at risk for bee venom hypersensitivity. Antibodies specific
CC for Api m 6 are useful for purifying the protein.
XX
SQ Sequence 67 AA;
Query Match 100.0%; Score 391; DB 23; Length 67;
Best Local Similarity 100.0%; Pred. No. 8.9e-32;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGFGLGGRGKCPNSIFSRCDRCQFCPNVVPKPLCIKICAPGCVCRGLYLRNKKKVC 60
DB 1 GGFGLGGRGKCPNSIFSRCDRCQFCPNVVPKPLCIKICAPGCVCRGLYLRNKKKVC 60
QY 61 VPRSKCG 67
DB 61 VPRSKCG 67
RESULT 2
RAY30434
ID RAY30434 standard; Protein; 61 AA.
XX
AC RAY30434;
XX
DT 15-NOV-1999 (first entry)
XX
DE Mature nematode extracted anticoagulant protein NamNAP.
XX
KW Nematode extracted anticoagulant protein; NAP; anticoagulant;
KW serine protease inhibitor; NAP domain; factor VIIa/TF.
XX
OS Necator americanus.
XX
PN US5955294-A.
XX
PD 21-SEP-1999.
XX
PF 19-APR-1996; 96US-0634641.
XX
PR 19-APR-1996; 96US-0634641.
PR 18-OCT-1994; 94US-0326110.
PR 05-JUN-1995; 95US-0461965.
PR 05-JUN-1995; 95US-0461965.
PR 05-JUN-1995; 95US-0465380.
PR 05-JUN-1995; 95US-0486397.
PR 05-JUN-1995; 95US-0486399.
PR 17-OCT-1995; 95WO-US13231.
XX
PA (CORV-) CORVAS INT INC.
XX
PI Bergum PW, Gansemans YGJ, Jespers LS, Laroche YR;
PI Lauwereys MJ, Messens JHL, Moyle M, Stanssens PEH;
PI Vlasuk GP;
XX
DR WPI; 1999-539569/45.
XX
XX Screening an isolated protein for Nematode-extracted Anticoagulant
PT Protein domains
PT
XX

PS Disclosure; Columns 143-144; 197pp; English.
XX
CC The present sequence represents a nematode extracted anticoagulant
CC protein (NAP). The protein has activity as an anticoagulant and/or serine
CC protease inhibitor. The protein contains at least one NAP domain which
CC has selective inhibitory activity for factor VIIa/TF. The specification
CC describes a method for screening an isolated protein at least one domain
CC for factor VIIa/TF selective inhibitory activity. The method comprises
CC determining the time to clotting effected by a concentration of the
CC isolated protein in an ex vivo prothrombin time (PT) assay and an ex vivo
CC activated partial thromboplastin time (APTT) assay; calculating
CC prolongation of clotting effected by the isolated protein in each of
CC the PT and APTT assay, with respect to a baseline clotting value for
CC each assay, where prolongation of clotting is calculated as fold
CC elevation of clotting time relative to a baseline clotting value, where
CC a doubling of clotting time is deemed a two-fold elevation; and
CC calculating a PT to APTT prolongation ratio, where a ratio at least
CC one is indicative of factor VIIa/TF inhibitory activity. The method is
CC useful for determining if a protein has factor VIIa/TF inhibitory
CC activity.
XX
SQ Sequence 61 AA;
Query Match 21.0%; Score 82; DB 20; Length 61;
Best Local Similarity 30.9%; Pred. No. 0.43;
Matches 17; Conservative 9; Mismatches 19; Indels 10; Gaps 2;
QY 12 CFSNEIFSRCDRCQFCPNVVPKPLCIKICAPGCV----CRGLYLRNKKKVCV 61
DB 4 CPANEEWRECGTFCPEKCNQMP-----DICTMNCIVDVCCKEGYKRHTKGCL 53
RESULT 3
AAB15319
ID AAB15319 standard; Protein; 61 AA.
XX
AC AAB15319;
XX
DT 19-DEC-2000 (first entry)
XX
DE N. americana nematode-extracted anticoagulant protein NamNAP.
XX
KW Nematode-extracted anticoagulant protein; NamNAP; blood clotting;
KW canine hookworm; thrombosis; vaccine.
XX
OS Necator americanus.
XX
PN US6087487-A.
XX
PD 11-JUL-2000.
XX
PF 12-FEB-1999; 99US-0249451.
XX
PR 17-OCT-1995; 95WO-US13231.
PR 17-APR-1997; 97US-0809455.
PR 18-OCT-1994; 94US-0326110.
PR 05-JUN-1995; 95US-0461965.
PR 05-JUN-1995; 95US-0465380.
PR 05-JUN-1995; 95US-0486397.
PR 05-JUN-1995; 95US-0486399.
XX
PA (CORV-) CORVAS INT INC.
XX
PI Lauwereys MJ, Stanssens PEH, Jespers LS, Gansemans YGJ, Moyle M;
PI Bergum PW, Messens JHL, LaRoche YR, Vlasuk GP;
XX
DR WPI; 2000-531359/48.
DR N-PSDB; AAB73383.
XX
XX New cDNA molecule encoding a protein having factor Xa inhibitory
PT activity for preventing and treating blood clotting disorders,
PT comprises nematode-extracted anticoagulant protein domains -
XX

CC The present sequence represents a nematode extracted anticoagulant
CC protein (NAP). The protein has activity as an anticoagulant and/or serine
CC protease inhibitor. The protein contains at least one NAP domain which
CC has selective inhibitory activity for factor VIIa/TF. The specification

The present invention describes a transgenic algal cell (I) of the genus *Chlamydomonas* comprising reproductive genetic material comprising a nucleotide sequence capable of expressing chicken type I Metallothionein. Also described is a method of removing metal from an aqueous medium containing at least one dissolved or suspended metal. The transgenic algae are used for the selective separation of metals, particularly the separation of precious and desirable metals such as gold and uranium, from other metals such as cadmium, zinc and copper. The method can be used to facilitate the selective recovery of precious and rare metals from mineral sources where aqueous media can be used, such as in natural surface water flows, ground water and where water may be introduced. The method is suitable for well-drilling, soil and water remediation arts, mining fields, and industrial engineering. The present sequence represents a Class I metallothionein

CC given in the present invention.

XX Sequence 59 AA;

SQ Query Match 16.2%; Score 63.5; DB 21; Length 59;

Best Local Similarity 35.2%; Pred. No. 29;

Matches 19; Conservative 2; Mismatches 18; Indels 15; Gaps 3;

QY 7 GSGKCPSEIFRCDRCORPCPNVVPKLCIKICAPGCVCLGYLRNKKVC 60

Db 15 GSGCKC-SNCACTSKKSCPCPCPSDK-----CASGCV-----KGKTC 53

RESULT 6

ABP32832

ID ABP32832 standard; Protein; 59 AA.

XX

AC ABP32832;

XX 09-JUL-2002 (first entry)

XX Human ORF1805 protein, SEQ ID NO:3610.

XX Human; ORF; open reading frame; ORFX; drug screening; diagnosis;
XX disease monitoring; cytokine; cell proliferation; cell differentiation;
XX immune modulation; haematopoiesis regulation; tissue growth;
XX angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;
XX thrombolytic; tumour inhibition; bodily characteristic; fertility;
XX behaviour; cancer; proliferative disorder; neurological disorder;
XX cardiovascular disease; immune system disorder; organ transplantation;
XX tissue growth disorder; tissue regeneration disorder; diabetes mellitus;
XX hypothyroidism; cholesterol ester storage disease; infection; vulnery;
XX vasotropic; antipsoriatic; antidiabetic; cytostatic; nootropic;
XX neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;
XX cardiant; hypotensive; antithyroid; antiinflammatory; immunomodulator;
XX dermatological; analgesic; virucide; antibacterial; fungicide.

XX Homo sapiens.

XX WO200190366-A2.

XX 29-NOV-2001.

XX 24-MAY-2001; 2001WO-US17076.

XX 24-MAY-2000; 2000US-206690P.

XX (CURA-) CURAGEN CORP.

XX Leach MD, Shinkets RA;

XX WPI; 2002-106200/14.

XX N-PSDB; ABN76858.

XX Novel human polypeptides and polynucleotides useful for diagnosing,
XX preventing and treating cardiovascular disease, neurodegenerative,
XX hyperproliferative disorders and disorders related to organ
XX transplantation -
XX Claim 10; Page 1162; 2508pp; English.

XX Sequences ABP31028-ABP35561 represent 4534 novel human proteins
XX designated ORF (open reading frame) 1-4534, and sequences ABN75054-
XX ABN79587 represent cDNAs encoding them. The invention also encompasses
XX polypeptides at least 80% identical to the ORF1-ORF4534 (collectively
XX referred to as ORFX) proteins, polynucleotides at least 85% identical to
XX the ORFX nucleic acid sequences, vectors and host cells comprising ORFX
XX polynucleotides, the recombinant production of ORFX proteins, antibodies
XX specific for ORFX proteins, methods of detecting ORFX polynucleotides and
XX polypeptides, methods of screening for modulators of ORFX expression or
XX activity, and methods of screening individuals for a predisposition to an
XX ORFX-associated disorder. The ORFX proteins of the invention have a wide
XX range of biological activities, such as cytokine, cell proliferation,

CC cell differentiation, immune modulation, haematopoiesis regulation,
CC tissue growth, angiogenesis, activin or inhibin activity, chemotactic/
CC chemokinetic activity, haemostatic activity, thrombolytic activity,
CC receptor/ligand, antiinflammatory activity, tumour inhibition activity,
CC and antifective activity, and may also be involved in the determination
CC of bodily characteristics, fertility and behaviour. ORFX proteins,
CC nucleic acids and antibodies may be used in the treatment of cancers,
CC other proliferative disorders such as psoriasis and benign tumours,
CC neurological disorders such as epilepsy and Alzheimer's disease,
CC cardiovascular diseases, immune system disorders, disorders related to
CC organ transplantation, disorders of tissue growth and regeneration,
CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester
CC storage disease, and infectious diseases caused by viral, bacterial,
CC fungal and other pathogens. ORFX nucleic acids may also be used as a
CC source of primers and probes, in the detection of ORFX genomic sequences
CC or transcripts, in the identification and cloning of homologous
CC sequences, in genetic diagnosis, and in forensic biology. The ORFX
CC nucleic acids may additionally be used to produce transgenic animals
CC which may be useful for studying the function and/or activity of ORFX
CC protein, and in drug screening. The ORFX proteins may also be used as
CC immunogens to generate specific antibodies, which are useful in the
CC diagnosis, treatment and monitoring of ORFX-associated diseases.

XX SQ Sequence 59 AA;

Query Match 16.2%; Score 63.5; DB 23; Length 59;

Best Local Similarity 28.9%; Pred. No. 29;

Matches 13; Conservative 7; Mismatches 12; Indels 13; Gaps 2;

QY 25 QCRFCPNV---VPKPLCIKICAPGCVCLGYLRNKKVCVPSK 66

Db 5 CAAVCPVCAAVCLPVCVQLCVSMVC-----LCMHATMC 39

RESULT 7

ABB77804

ID ABB77804 standard; Protein; 62 AA.

XX ABB77804;

XX 27-SEP-2002 (first entry)

XX Consensus sequence of mammalian metallothionein.

XX Streptavidin; metallothionein; heavy metal ion; tumour cell imaging;

XX biotin; radiotherapy; mass spectrometry.

XX Mammalia.

XX Key Location/Qualifiers

XX Domain 1..29 /note= "beta domain"

XX Domain 30..62 /note= "alpha domain"

XX US6391590-B1.

XX 21-MAY-2002.

XX 21-OCT-1991; 91US-0780717.

XX 21-OCT-1991; 91US-0780717.

XX (REGC) UNIV CALIFORNIA.

XX Sano T, Glazer AM, Cantor CR;

XX WPI; 2002-488386/52.

XX A recombinant streptavidin-metallothionein chimeric protein is useful
XX to add or remove heavy metal ions into biotin-containing biological
XX material, particularly for tumour imaging, radiotherapy, and DNA or
XX protein labelling -

XX PS Disclosure; Column 3; 9pp; English.

XX CC The present sequence represents mammalian metallothionein. It is

XX CC used to produce chimeric proteins of the invention. The specification

XX CC describes a recombinant bifunctional streptavidin-metallothionein

XX CC chimeric protein. This protein is produced by introducing into a host

XX CC cell nucleic acid encoding a bifunctional fusion protein having a

XX CC streptavidin and a metallothionein moiety, and incubating the cell to

XX CC express the fusion protein. The streptavidin moiety consists of residues

XX CC 16-133 of mature streptavidin. The chimeric protein is used to

XX CC incorporate heavy metal ions into biological materials containing biotin,

XX CC or to remove heavy metal ions from the biological material. Specific uses

XX CC include loading cancerous tissue with heavy metal ions for imaging of

XX CC tumour cells and radiotherapy, and labeling DNA and proteins for

XX CC detection on gels or blots by surface scanning mass spectrometry.

XX SQ Sequence 62 AA;

Query Match 16.0%; Score 62.5; DB 23; Length 62;

Best Local Similarity 37.3%; Pred. No. 38;

Matches 19; Conservative 2; Mismatches 23; Indels 7; Gaps 3;

Qy 1 GGFGLGGRGKCPNEIFSRDGRQRFQPCPNVFPKPLCIKICAPGCVCLRG 51

Db 10 GGSCTCAGSKCKECKCTS-CKKSCSCP-----VGCAK-CAQGCVCCKAG 53

RESULT 8

AAO21321

ID AAO21321 standard; Protein; 63 AA.

XX AC AAO21321;

XX DT 05-AUG-2002 (first entry)

XX DE Arabidopsis thaliana KCP-like protein, SEQ ID NO 76.

XX KW Antimicrobial; transgenic; plant; potato snakin antimicrobial protein;

XX KW GAS44; GAS45; GST1 homologue; lysine- and cysteine- rich peptide;

XX KW KCP-like polypeptide; modulating; disease resistance.

XX OS Arabidopsis thaliana.

XX PN WO200222821-A2.

XX PD 21-MAR-2002.

XX PF 13-SEP-2001; 2001WO-US28429.

XX PR 13-SEP-2000; 2000US-232569P.

XX PR 11-SEP-2001; 2001US-0950933.

XX PA (PION-) PIONEER HI-BRED INT INC.

XX PI Simmons CR, Navarro Acevedo PA;

XX DR WPI; 2002-425842/45.

XX PT New polynucleotide encoding lysine- and cysteine-rich peptides-like

XX PT polypeptide useful for modulating the polypeptide level in a plant

XX PT cell, enhancing disease resistance -

XX PS Disclosure; Page 154; 163pp; English.

XX CC The invention relates to an isolated polynucleotide encoding a

XX CC polypeptide which is related to potato snakin antimicrobial protein and

XX CC GAS44 or GAS45 or GST1 homologues, which is referred to as lysine- and

XX CC cysteine- rich peptides (KCP)-like polypeptide, having a nucleotide

XX CC sequence from 36 sequences of defined base pairs, given in the

XX CC specification. A recombinant expression cassette comprising the isolated

XX CC polynucleotide of the invention is useful for modulating the level of

XX CC (KCP)-like polypeptides in a plant cell, where the level of (KCP)-like

XX CC polypeptides is increased, and disease resistance is enhanced. This

XX CC sequence represents a protein relating to the KCP-like proteins of the

XX CC invention.

XX SQ Sequence 63 AA;

Query Match 16.0%; Score 62.5; DB 23; Length 63;

Best Local Similarity 37.8%; Pred. No. 39;

Matches 17; Conservative 4; Mismatches 19; Indels 5; Gaps 3;

Qy 20 RCDGRQRFQPCNV-VPKP---LCIKICAPGCVCLRGVLRNKKKVC 60

Db 4 QCGGQCTTRCSNTKYRPFQCFQCKCAK-CLCVPPGTGNGKQVC 47

RESULT 9

AAB60286

ID AAB60286 standard; Protein; 43 AA.

XX AC AAB60286;

XX DT 30-MAR-2001 (first entry)

XX DE Human factor IX (hFIX) exon-encoded fragment, SEQ ID NO:9.

XX KW Age-related gene regulation; liver-specific; gene expression;

XX KW human factor IX; hFIX; AEs'; AE3'; age-regulatable expression construct;

XX KW anticense therapy; gene therapy; thrombosis; cardiovascular disease;

XX KW diabetes; Alzheimer's disease; Parkinson's disease; cancer; osteoporosis;

XX KW osteoarthritis; dementia.

XX OS Homo sapiens.

XX PN WO200075279-A2.

XX PD 14-DEC-2000.

XX PF 06-JUN-2000; 2000WO-US15728.

XX PR 09-JUN-1999; 99US-0328925.

XX PA (UNWI) UNIV MICHIGAN.

XX PI Kurachi K, Kurachi S;

XX DR WPI; 2001-061708/07.

XX DR N-ESDB; AAF54018.

XX PT New regulatory elements that control age-related gene expression,

XX PT useful in gene therapy and for reducing Factor IX expression -

XX PS Disclosure; Fig 8C; 225pp; English.

XX CC The invention relates to nucleic acid sequences which regulate gene

XX CC expression in an age-related manner and/or in a liver-specific manner.

XX CC The invention identifies regions of the human factor IX (hFIX) gene, and

XX CC a region of the human protein C (hPC) gene, which are age-related

XX CC regulatory sequences. The hFIX age-related regulatory sequences are

XX CC designated AE5' (AAF54016) and AE3' (AAF54017) and are found in the 5'

XX CC UTR (at position 2164-2165 of AAF54018) and 3' UTR (at position

XX CC 34383-35655 of AAF54018) respectively. These elements act synergistically

XX CC to increase hFIX levels over the lifespan of an individual; however, they

XX CC can independently exert effects on hFIX mRNA in an age-related manner,

XX CC with AE5' acting to stabilise hFIX mRNA, and AE3' acting to increase hFIX

XX CC mRNA levels, over time. AE5' also directs liver-specific expression. The

XX CC hPC gene age-related regulatory sequence is found in the 5' UTR

XX CC (AAF54081), and contains two PEA-3 (polyoma virus activator 3) elements

XX CC 5'-GAGGAAA-3' and 5'-CAGGAG-3'. The age-related regulatory sequences of

XX CC the invention, along with their homologues, variants and fragments, may

XX CC be used in the construction of recombinant expression vectors for the

XX CC expression of a desired sequence in an age-related fashion in a host

XX CC cell. Preferred target genes for expression in such age-regulatable

XX CC expression vectors include those encoding proteins involved in blood

CC coagulation (e.g., the pro-coagulants factor IX and factor VIII, and the
 CC anti-coagulants protein C and antithrombin III), human
 CC alpha-1-antitrypsin, PEA-3 protein and reporter proteins such as
 CC luciferase. Preferred promoters for use in such age-regulatable
 CC expression vectors include the human factor IX promoter, the T7 promoter,
 CC the T3 promoter and the SP6 promoter. The expression vectors of the
 CC invention may be used in gene therapy to provide age-related and/or
 CC liver-specific expression of target genes. Age-regulatable constructs may
 CC be used in the treatment of such age-related conditions such as
 CC thrombosis, cardiovascular disease, diabetes, Alzheimer's disease,
 CC Parkinson's disease, cancer, osteoporosis, osteoarthritis and dementia.
 CC Specifically, they may be used to express factor IX antisense mRNA in the
 CC treatment of thrombotic conditions associated with the natural
 CC age-related rise in factor IX expression. Transgenic cells or animals
 CC that contain vectors of the invention are useful as models of these
 CC diseases, in screening for potential therapeutic agents and for studying
 CC normal processes such as ageing and gene expression. Fragments and
 CC homologues of age-related regulatory sequences, are useful as probes to
 CC detect, isolate or identify other such sequences in samples. The present
 CC sequence represents a fragment of hFIX.

XX SQ Sequence 43 AA;

Query Match 15.9%; Score 62; DB 22; Length 43;
 Best Local Similarity 33.3%; Pred. No. 31;
 Matches 14; Conservative 5; Mismatches 15; Indels 8; Gaps 2;

Qy 22 DGRQRCFCNVVPLCLIKICAPGVCPLGY-LRNKKVCVP 62
 Db 8 NGRCEQFCNSADNKV-----CSTEGYRLAENQKSCFP 42

RESULT 10

AAV57822
 ID AAY57822 standard; protein; 61 AA.

XX AC AAY57822;

XX DT 22-MAR-2000 (first entry)

XX DE Rabbit liver metallothionein Class II amino acid sequence.

XX KW Metallothionein; metal recovery; remediation; heavy metal;
 XX KW precious metal; phytochelatin; green algae; Chlamydomonas reinhardtii.
 XX OS Oryctolagus cuniculus.

XX PN WO9960838-A1.

XX PD 02-DEC-1999.

XX PF 28-MAY-1999; 99WO-US12007.

XX PR 28-MAY-1998; 98US-0087374.

XX PA (OHIS) UNIV OHIO STATE RES FOUND.

XX PI Sayre RT, Traina SJ;

XX DR WPI; 2000-086646/07.

XX FT Novel method for metal recovery, remediation and separation -

XX PS Disclosure; Fig 1; 86pp; English.

XX SQ The present invention describes a transgenic algal cell (I) of the
 CC Genus Chlamydomonas comprising reproductive genetic material comprising
 CC a nucleotide sequence capable of expressing chicken type I
 CC Metallothionein. Also described is a method of removing metal from
 CC an aqueous medium containing at least one dissolved or suspended
 CC metal. The transgenic algae are used for the selective separation of
 CC metals, particularly the separation of precious and desirable metals
 CC such as gold and uranium, from other metals such as cadmium, zinc and

CC copper. The method can be used to facilitate the selective recovery of
 CC precious and rare metals from mineral sources where aqueous media can
 CC be used, such as in natural surface water flows, ground water and where
 CC water may be introduced. The method is suitable for well-drilling,
 CC soil and water remediation arts, mining fields, and industrial
 CC engineering. The present sequence represents a Class II metallothionein
 CC given in the present invention.

XX SQ Sequence 61 AA;

Query Match 15.9%; Score 62; DB 21; Length 61;
 Best Local Similarity 38.5%; Pred. No. 42;
 Matches 15; Conservative 2; Mismatches 10; Indels 12; Gaps 3;

Qy 11 KCPSENEIFSCDGRQRCFCNVVPLCLIKICAPGVCVR 49
 Db 25 KCTS-----CKKSCCCP-----PGCAK-CAGGCLCK 51

RESULT 11

AAB61455
 ID AAB61455 standard; protein; 67 AA.

XX AC AAB61455;

XX DT 04-APR-2001 (first entry)

XX DE Metallothionein domain consensus.

XX KW TANGO; MANGO; colon; inflammation; tumor; renal; liver; lung;
 XX KW autoimmune; allergy; cardiovascular; brain; degenerative; placental;
 XX KW pancreatic; skeletal; muscle.

XX OS Synthetic.

XX PN WO200100672-A1.

XX PD 04-JAN-2001.

XX PF 29-JUN-2000; 2000WO-US18184.

XX PR 29-JUN-1999; 99US-0342687.

XX PA (MILL-) MILLENNIUM PHARM INC.

XX PI Holtzman DA, Barnes TM, Fraser CC, Sharp JD;

XX DR WPI; 2001-050127/06.

XX PT Isolated cDNAs encoding TANGO 244, TANGO 246, TANGO 275, TANGO 300 and
 PT MANGO 245 proteins, useful in the treatment of inflammatory diseases
 PT (e.g. idiopathic ulcerative colitis), tumors, renal disorders and liver
 PT disorders (e.g. jaundice) -

XX PS Disclosure; Fig 13; 262pp; English.

XX SQ The present invention relates to cDNAs encoding TANGO 244,
 CC TANGO 246, TANGO 275, TANGO 300 and MANGO 245 proteins.

CC The nucleic acids, proteins and protein modulators are useful for
 CC treating colonic disorders, inflammatory diseases, tumors,
 CC renal disorders, liver disorders, lung disorders, autoimmune diseases,
 CC allergic diseases, cardiovascular diseases, brain disorders,
 CC degenerative diseases placental, pancreatic, skeletal and muscle
 CC disorders.

XX SQ Sequence 67 AA;

Query Match 15.7%; Score 61.5; DB 22; Length 67;
 Best Local Similarity 31.7%; Pred. No. 51;
 Matches 19; Conservative 5; Mismatches 29; Indels 7; Gaps 2;

Qy 1 GGFGLGGKGCPSNEIFSCDGRQRCFCNVVPLCLIKICAPGVCRLGRLNKKVC 60
 ||| : : : : : ||| : : : : :
 ||| : : : : : ||| : : : : :
 ||| : : : : : ||| : : : : :

Db 11 GGSCTGTSCKCNKCKTS-CKKSCCSCPAGCSK-----CAGGCVCKGGAASETSKC 63

```
RESULT 12
AAV04648
ID AAV04648 standard, peptide; 47 AA.
XX
AC AAV04648;
XX
DT 22-JUN-1999 (first entry)
XX
DE Factor IXa catalytic and interacting domains.
XX
KW Receptor; catalytic domain; Factor IXa; Factor Xa; tissue factor; angina;
KW blood clotting disorder; thrombosis; restenosis; myocardial infarction;
KW reclosure; cerebrovascular disease; hypercoagulability; anticoagulant;
KW peripheral arterial occlusive disease; pulmonary embolism; cyclic.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Disulfide-bond 8..19
FT Domain 10..12
FT /note= "Factor VIIa interacting domain"
FT Disulfide-bond 13..29
FT Domain 14..16
FT /note= "catalytic domain"
FT Disulfide-bond 31..44
XX
FN WO9913062-A1.
XX
PD 18-MAR-1999.
XX
PF 08-SEP-1998; 98WO-CB02700.
XX
PR 09-SEP-1997; 97GB-0019157.
XX
PA (MATT/) MATTHEWS D P.
PA (NYCO-) NYCOMED IMAGING AS.
XX
PI Fischer PM, Sakariassen KS;
XX
WPI; 1999-215061/18.
XX
New anticoagulant compounds
XX
Disclosure; Fig 1; 50pp; English.
XX
Peptides AAY04625-Y04647 represent claimed compounds which are capable
of interacting with an internal receptor in the catalytic domain of
Factor IXa (FIXa) or Factor X (FX) defined by the residues Ile290,
Ala291, Asp292, Tyr293, Thr294, Glu374 and Phe378 of FIXa, and Leu300,
Pro301, Glu302, Trp305, Ala306, Lys385 and Phe389 of FXa, or the ligand
defined by residues Cys95-Cys99 of FIXa or Cys96-Cys100 of FXa. The
compounds can be used to prevent the formation of a functional
FVIIa/FIXa or FVIIa/FXa complex, so can be used to combat or prevent
blood clotting disorders, e.g. thrombosis (particularly vascular
thrombosis or deep vein thrombosis), acute myocardial infarction,
restenosis, angina, reclosure, cerebrovascular disease, peripheral
arterial occlusive disease, hypercoagulability or pulmonary embolism.
They can also be used to prevent the occurrence of blood clotting
disorders caused by e.g. grafting surgery, vessel wall potency
restoration or sepsis.
```

Query Match 15.6%; Score 61; DB 20; Length 47;

Best Local Similarity 34.1%; Pred. No. 42;

Matches 14; Conservative 4; Mismatches 15; Indels 8; Gaps 2;

Qy 23 GRCQRFQNVVVKPLCIKICAPGCVCLGY-LRNKKKVCVP 62

Db 13 GRCQFCKNSADNKKV-----CSCTEGYRLAENQKSCPE 46

RESULT 13

AA44778

ID AAB44778 standard; Protein; 56 AA.

XX

AC AAB44778;

XX

DT 12-FEB-2001 (first entry)

XX

DE Human secreted protein sequence encoded by gene 17 SEQ ID NO:77.

XX

OS Homo sapiens.

XX

FN WO200058336-A1.

XX

PD 05-OCT-2000.

XX

PF 23-MAR-2000; 2000WO-US07726.

XX

PR 26-MAR-1999; 99US-0126597.

XX

PR 07-JAN-2000; 2000US-0174877.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PI Rosen CA, Ruben SM, Komatsoulis G;

XX

DR WPI; 2000-602355/57.

XX

DR N-PSDB; AAC79815.

XX

XX

PT Nucleic acid encoding human secreted proteins, used to treat, prevent,

XX

PT ameliorate or diagnose medical conditions such as cancer, and

XX

PT autoimmune diseases -

XX

PS Claim 11; Page 362; 391pp; English.

XX

CC The polynucleotide sequences given in AAC79799 to AAC79848 encode the

XX

CC human secreted proteins given in AAB44762 to AAB44811. AAB44812 to

XX

CC AAB44829 represent human secreted polypeptide sequences and proteins

XX

CC homologous to them, which are used in the exemplification of the present

XX

CC invention. Human secreted proteins have activities based on the tissues

XX

CC and cells the genes are expressed in. Examples of activities are:

XX

CC immunosuppressive; antiarthritic; vasotropic; antineoplastic; antiproliferative;

XX

CC cytoskeletal; cardiac; vasotropic; cerebroprotective; neurotrophic;

XX

CC neuroprotective; antibacterial; virucide; fungicide; and

XX

CC ophthalmological. The polynucleotides and polypeptides can be used to

XX

CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,

XX

CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used

XX

CC in diagnosing a pathological condition or susceptibility to a

XX

CC pathological condition. Disorders which are diagnosed or treated include

XX

CC autoimmune diseases, hyperproliferative disorders, cardiovascular

XX

CC disorders, cerebrovascular disorders, angiogenesis, nervous system

XX

CC disorders, infections caused by bacteria, viruses and fungi and ocular

XX

CC disorders. The polypeptides can also be used to aid wound healing and

XX

CC epithelial cell proliferation, to prevent skin aging due to sunburn, to

XX

CC maintain organs before transplantation, for supporting cell culture of

XX

CC primary tissues, to regenerate tissues and in chemotaxis. The

XX

CC polypeptides can also be used as a food additive or preservative to

XX

CC increase or decrease storage capabilities. AAC79790 to AAC79798 and

XX

CC AAB44761 represent sequences used in the exemplification of the present

XX

CC invention.

XX

Sequence 56 AA;

Query Match 15.6%; Score 61; DB 21; Length 56;

Best Local Similarity 29.8%; Pred. No. 49;
Matches 14; Conservative 3; Mismatches 18; Indels 12; Gaps 2;

Qy 21 CDGRQRFPCNPVVPKPLIKICAPGCVCRIGYLRNKKVCVPRSKCG 67
 | | | | | |
Db 4 CVQACVHMC-----LCVHVCAASVHVCECAVH-----CVRECAAG 38
 | | | | |

RESULT 14
AAU45263

ID	AAU45263 standard; Protein; 66 AA.
1	MA
2	MA
3	MA
4	MA
5	MA
6	MA
7	MA
8	MA
9	MA
10	MA
11	MA
12	MA
13	MA
14	MA
15	MA
16	MA
17	MA
18	MA
19	MA
20	MA
21	MA
22	MA
23	MA
24	MA
25	MA
26	MA
27	MA
28	MA
29	MA
30	MA
31	MA
32	MA
33	MA
34	MA
35	MA
36	MA
37	MA
38	MA
39	MA
40	MA
41	MA
42	MA
43	MA
44	MA
45	MA
46	MA
47	MA
48	MA
49	MA
50	MA
51	MA
52	MA
53	MA
54	MA
55	MA
56	MA
57	MA
58	MA
59	MA
60	MA
61	MA
62	MA
63	MA
64	MA
65	MA
66	MA

AAU45263;

DT 27-FEB-2002 (first entry)

DE Propionibacterium acnes immunogenic protein #6159.

KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
 KW dermatological; osteopathic; neuroprotectant.

OS Propionibacterium acnes.

PN WO200181581-A2.

PD 01-NOV-2001.

20-APR-2001: 2001WO-US12865.

21-APR-2000: 2000US-199047P.

PR 02-JUN-2000; 2000US-208841P.
PR 07-JUL-2000; 2000US-216747P.

PA (CORI-) CORIXA CORP.

PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;

DR WPI; 2001-616774/71.

DR N-PSDB; AAS59525.

PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -

PS Example 1; SEQ ID No 6458; 1069pp; English.

Sequences AAU39105-AAU68017 represent *Propionibacterium* acnes immunogenic polypeptides. The proteins and their associated DNA sequences are used in the treatment, prevention and diagnosis of medical conditions caused by *P. acnes*. The disorders include SAPHO syndrome (synovitis, acne, pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis. *P. acnes* is also involved in infections of bone, joints and the central nervous system, however it is particularly involved in the inflammatory lesions associated with acne vulgaris. A method for detecting the presence or absence of *P. acnes* in a patient comprises contacting a sample with a binding agent that binds to the proteins of the invention and determining the amount of bound protein in the sample. The polypeptides may be used as antigens in the production of antibodies specific for *P. acnes* proteins. These antibodies can be used to downregulate expression and activity of *P. acnes* polypeptides and therefore treat *P. acnes* infections. The antibodies may also be used as diagnostic agents for determining *P. acnes* presence, for example, by enzyme linked immunosorbent assay (ELISA).

CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published pct sequences.

SQ , Sequence 66 AA:

Query Match

15.6%; Score 61; DB 22; Length 66:

Best Local Similarity 31.8%; Pred. No. 57;
Matches 21; Conservative 6; Mismatches 27; Indels 12; Gaps 4;

Qy 2 GFGGLGGRKC--PSNEIFSRCDGRCQFCFNNVXPPLCIKTCAPGCVCRGLGLRNKKK 58

Dd 7 GRGELISACKGGRFSDQ--RRCSAQCGRLLGSRPVSRCVR-----CIC--CSIRASPL 57

Qy 59 VCVPRS 64

Db 58 PCQPPS 63

RESULT 15

ABF01364
ID ABP01364 standard: Protein: 60 AA

AC ABP01364:

25-JUN-2002 (first entry)

AA
DE
Human ORFX protein sequence SEO ID NO:2710.

Human; open reading frame; ORF; gene therapy; cancer; cirrhosis;
 KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
 KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
 KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
 KW hypertension; hypothyroidism; cholesterol ester storage disease;
 KW immune deficiency; immune disorder; infectious disease;
 KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
 KW myasthenia gravis.

OS Homo sapiens.

PN WO200192523-A2.

06-DEC-2001.

29-MAY-2001:

PR 30-MAY-2000: 2000US-206132P.

XX
FK Z9-AUG-2007; 200005-227/16P.

FA (CURA-) CURAGEN CORP. XX

XX
XX
IT IS A FACT THAT
TEACHING IS

DR WF1; 2002-106308/14.
DR N-PSDB; ABN17116.

PT Novel human polyp

PT hyperproliferative disorders and autoimmune disorders - preventing and treating cardiovascular disease, neurodegenerative,

PS Disclosure; SEQ ID 2710; 1037pp; English.

The present invention describes substantially purified human proteins (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1 in the specification). ABN15162 to ABN27252 encode the human ORFX proteins given in ABP00010 to ABP1500. ORFX proteins are useful for treating or preventing a pathology associated with an ORFX-associated disorder in humans, and in the manufacture of a medicament for treating a syndrome associated with ORFX-associated disorder. ORFX polynucleotide sequences can be used in gene therapy. ORFX sequences can be used in the treatment of cancer, hyperproliferative disorders, cirrhosis of liver, psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage, osteoarthritis, neurodegenerative disorders, disorders related to organ transplantation, cardiovascular diseases, diabetes mellitus, systemic lupus erythematosus, hypertension, hypothyroidism, cholesterol ester storage disease, various immune deficiencies and disorders, infectious diseases, autoimmune disorders such as multiple sclerosis, rheumatoid arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. ORFX proteins are also useful for treating burns, incisions, ulcers, for treating osteoporosis.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2003, 08:16:59 ; Search time 29 Seconds
(without alignments)
421.775 Million cell updates/sec

Title: US-09-506-978-1

Perfect score: 391

Sequence: 1 GGFGLGGRGKCPSPNEIFSR.....CRGLYLRNKKKVCVPRSKCG 67

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues

Total number of hits satisfying chosen parameters: 208767

Minimum DB seq length: 0

Maximum DB seq length: 67

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA.*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pdb.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pdb.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pdb.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pdb.*
5: /cgn2_6/ptodata/2/pubpaa/PCTUS_NEW_PUB.pdb.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pdb.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pdb.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pdb.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pdb.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pdb.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pdb.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pdb.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pdb.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pdb.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pdb.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pdb.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pdb.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pdb.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	391	100.0	67	12	US-10-174-151-1
2	119	30.4	67	12	US-10-087-887-40
3	82	21.0	61	11	US-09-498-272-61
4	68	17.4	58	11	US-09-498-272-60
5	63	16.1	60	9	US-09-790-264-59
6	63	16.1	60	15	US-10-269-353-59
7	62.5	16.0	63	10	US-09-950-933A-76
8	59.5	15.2	61	11	US-09-919-039-195
9	57.5	14.7	61	10	US-09-981-353-115
10	57.5	14.7	61	11	US-09-919-039-245
11	57	14.6	47	12	US-10-087-887-41
12	56.5	14.5	61	10	US-09-981-353-120
13	56.5	14.5	61	11	US-09-919-039-31
14	56.5	14.5	61	11	US-09-919-039-272
15	55.5	14.2	38	9	US-09-847-185-47
					Sequence 1, Appl
					Sequence 40, Appl
					Sequence 61, Appl
					Sequence 60, Appl
					Sequence 59, Appl
					Sequence 76, Appl
					Sequence 195, App
					Sequence 115, App
					Sequence 245, App
					Sequence 41, Appl
					Sequence 120, App
					Sequence 31, Appl
					Sequence 272, App
					Sequence 47, Appl

16	55.5	14.2	38	15	US-10-224-286-47	Sequence 47, Appl
17	55.5	14.2	45	9	US-09-817-647-11	Sequence 11, Appl
18	55.5	14.2	45	10	US-09-877-665-11	Sequence 11, Appl
19	55.5	14.2	45	14	US-10-136-573A-11	Sequence 11, Appl
20	55.5	14.2	45	15	US-10-215-862-11	Sequence 12, Appl
21	55.5	14.2	46	14	US-10-201-945-12	Sequence 12, Appl
22	55.5	14.2	58	15	US-10-231-778-229	Sequence 229, App
23	54.5	13.9	58	9	US-09-865-578-11	Sequence 11, Appl
24	54	13.8	44	15	US-10-072-602B-490	Sequence 490, App
25	53	13.6	38	11	US-09-764-891-2837	Sequence 2837, App
26	53	13.6	46	9	US-09-864-761-34995	Sequence 34995, A
27	52	13.3	42	15	US-10-252-340-15	Sequence 15, Appl
28	52	13.3	43	12	US-09-962-756-1501	Sequence 1501, App
29	51	13.0	57	12	US-10-029-386-28520	Sequence 28520, A
30	51	13.0	58	12	US-09-974-026-4	Sequence 4, Appl
31	51	13.0	64	12	US-10-160-162-165	Sequence 165, App
32	51	13.0	64	12	US-09-820-649-165	Sequence 165, App
33	51	13.0	67	12	US-09-933-767-608	Sequence 608, App
34	51	13.0	67	15	US-10-023-282-608	Sequence 608, App
35	50.5	12.9	57	15	US-10-231-778-233	Sequence 233, App
36	50	12.8	37	12	US-09-962-756-1460	Sequence 1460, App
37	50	12.8	43	12	US-09-962-756-1437	Sequence 1437, App
38	50	12.8	43	12	US-09-962-756-1472	Sequence 1472, App
39	50	12.8	43	12	US-09-962-756-1476	Sequence 1476, App
40	50	12.8	67	11	US-09-764-891-3803	Sequence 3803, App
41	49.5	12.7	57	12	US-09-974-026-20	Sequence 20, Appl
42	49.5	12.7	58	12	US-10-167-351-66	Sequence 66, Appl
43	49.5	12.7	58	12	US-10-038-722-56	Sequence 56, Appl
44	49.5	12.7	67	12	US-10-038-722-90	Sequence 90, Appl
45	49	12.5	40	9	US-09-917-340-73	Sequence 73, Appl

ALIGNMENTS

RESULT 1

US-10-174-151-1
; Sequence 1, Application US/10174151
; Publication No. US20030165514A1
; GENERAL INFORMATION:
; APPLICANT: Spertini, Francois
; TITLE OF INVENTION: NOVEL BEE VENOM POLYPEPTIDES AND METHODS OF USE THEREOF
; FILE REFERENCE: 18519-001
; CURRENT APPLICATION NUMBER: US/10/174,151
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US/09/506,978
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 67
; TYPE: PRT
; ORGANISM: Apis mellifera
US-10-174-151-1

Query Match 100.0%; Score 391; DB 12; Length 67;
Best Local Similarity 100.0%; Pred. No. 2.3e-34;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	GGFGLGGRGKCPSPNEIFSRCDGRQFCNVVPKLCIKICAPGCVCLYLRNKKVC	60
Db	1	GGFGLGGRGKCPSPNEIFSRCDGRQFCNVVPKLCIKICAPGCVCLYLRNKKVC	60
Qy	61	VPRSKCG	67
Db	61	VPRSKCG	67

RESULT 2

US-10-087-887-40
; Sequence 40, Application US/10087887
; Publication No. US20030198957A1
; GENERAL INFORMATION:


```

; PRIOR FILING DATE: 1998-04-23
; PRIOR APPLICATION NUMBER: US 09/298,531
; PRIOR FILING DATE: 1999-04-23
; PRIOR APPLICATION NUMBER: US 09/065,363
; PRIOR FILING DATE: 1998-04-23
; PRIOR APPLICATION NUMBER: US 09/337,930
; PRIOR FILING DATE: 1999-06-22
; PRIOR APPLICATION NUMBER: US 09/102,705
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: US 09/363,630
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: US 09/124,538
; PRIOR FILING DATE: 1998-07-29
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 59
; LENGTH: 60
; TYPE: PRT
; ORGANISM: Rattus norvegicus
; US-09-790-264-59

Query Match          16.1%; Score 63; DB 9; Length 60;
Best Local Similarity 38.1%; Pred.No.11;
Matches 16; Conservative 2; Mismatches 20; Indels 4; Gaps 2;

Qy 10 GKCPSEIFSRCDGRQRCFQPNVVPKPLCIKICAPGC--VCR 49
      |||||
Db 17 GKCPKNP--PRSIGTCVELCSGDSQSCPNIQKCCSNGCGHVK 56

RESULT 6
US-10-269-353-59
; Sequence 59, Application US/10269353
; Publication No. US2003010447A1
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Goodearl, Andrew D.J.
; APPLICANT: McCarthy, Sean A.
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
; TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OT
; TITLE OF INVENTION: USES
; FILE REFERENCE: MPI2000-5380MNCINIM
; CURRENT APPLICATION NUMBER: US/10/269,353
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: US 09/790,264
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: US 09/065,661
; PRIOR FILING DATE: 1998-04-23
; PRIOR APPLICATION NUMBER: US 09/298,531
; PRIOR FILING DATE: 1999-04-23
; PRIOR APPLICATION NUMBER: US 09/065,363
; PRIOR FILING DATE: 1998-04-23
; PRIOR APPLICATION NUMBER: US 09/337,930
; PRIOR FILING DATE: 1999-06-22
; PRIOR APPLICATION NUMBER: US 09/102,705
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: US 09/363,630
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: US 09/124,538
; PRIOR FILING DATE: 1998-07-29
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 59
; LENGTH: 60
; TYPE: PRT
; ORGANISM: Rattus norvegicus
; US-10-269-353-59

Query Match          16.1%; Score 63; DB 15; Length 60;
Best Local Similarity 38.1%; Pred.No.11;
Matches 16; Conservative 2; Mismatches 20; Indels 4; Gaps 2;

Qy 10 GKCPSEIFSRCDGRQRCFQPNVVPKPLCIKICAPGC--VCR 49

```



```

; TITLE OF INVENTION: PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-285
; CURRENT APPLICATION NUMBER: US/10/087,887
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: 60/273,049
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: 60/279,883
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/277,791
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/281,248
; PRIOR FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: 60/282,864
; PRIOR FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: 60/282,537
; PRIOR FILING DATE: 2001-04-09
; PRIOR APPLICATION NUMBER: 60/282,867
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 104
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 41
; LENGTH: 47
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-087-887-41

Query Match      14.6%; Score 57; DB 12; Length 47;
Best Local Similarity 23.4%; Pred. No. 37;
Matches 11; Conservative 9; Mismatches 13; Indels 14; Gaps 1;

QY 34 KPPLCIKICAPG-----CVCRLGYLRNKKKVCVPRSKC 66
Db 1 PGPDGLDVATCHEHATCQREGKICICNYGFVNGRTQCVDNKNC 47

RESULT 12
US-09-981-353-120
; Sequence 120, Application US/09981353
; Patent No. US20020160382A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy W.
; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
; FILE REFERENCE: PA-0038 US
; CURRENT APPLICATION NUMBER: US/09/981,353
; CURRENT FILING DATE: 2001-10-11
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 120
; LENGTH: 61
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20020160382A1 2757583CD1
US-09-981-353-120

Query Match      14.5%; Score 56.5; DB 10; Length 61;
Best Local Similarity 37.5%; Pred. No. 53;
Matches 15; Conservative 5; Mismatches 15; Indels 5; Gaps 4;

QY 10 GKCPSEIFSRCDGRCQRCFPCPNVVPKPLCIKICAPGCVCR 49
Db 17 GSCKCKE--CKCTS--CKKSCCSCPVG--CSK-CAQGCVCCK 51

RESULT 13
US-09-919-039-31
; Sequence 31, Application US/09919039
; Publication No. US20030108871A1
; GENERAL INFORMATION:
; APPLICANT: Kaser, Matthew R.
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES
```

```

; FILE REFERENCE: PA-0035 US
; CURRENT APPLICATION NUMBER: US/09/919,039
; CURRENT FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: 60/222,113
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 401
; SOFTWARE: PERL Program
; SEQ ID NO 31
; LENGTH: 61
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030108871A1 3860413CD1
US-09-919-039-31

Query Match      14.5%; Score 56.5; DB 11; Length 61;
Best Local Similarity 35.0%; Pred. No. 53;
Matches 14; Conservative 6; Mismatches 15; Indels 5; Gaps 4;

QY 10 GKCPSEIFSRCDGRCQRCFPCPNVVPKPLCIKICAPGCVCR 49
Db 17 GSCKCKE--CKCTS--CKKSCCSCPVG--CAK-CAQGCVCCK 51

RESULT 14
US-09-919-039-272
; Sequence 272, Application US/09919039
; Publication No. US20030108871A1
; GENERAL INFORMATION:
; APPLICANT: Kaser, Matthew R.
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES
; FILE REFERENCE: PA-0035 US
; CURRENT APPLICATION NUMBER: US/09/919,039
; CURRENT FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: 60/222,113
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 401
; SOFTWARE: PERL Program
; SEQ ID NO 272
; LENGTH: 61
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030108871A1 2757583CD1
US-09-919-039-272

Query Match      14.5%; Score 56.5; DB 11; Length 61;
Best Local Similarity 37.5%; Pred. No. 53;
Matches 15; Conservative 5; Mismatches 15; Indels 5; Gaps 4;

QY 10 GKCPSEIFSRCDGRCQRCFPCPNVVPKPLCIKICAPGCVCR 49
Db 17 GSCKCKE--CKCTS--CKKSCCSCPVG--CSK-CAQGCVCCK 51

RESULT 15
US-09-847-185-47
; Sequence 47, Application US/09847185
; Patent No. US20020076392A1
; GENERAL INFORMATION:
; APPLICANT: Soc Hoo, William
; TITLE OF INVENTION: MEMBRANE-BOUND CYTOKINE COMPOSITIONS
; COMPRISING GM-CSF AND METHODS OF MODULATING AN IMMUNE
; RESPONSE USING SAME
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CAMPBELL & FLORES, LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
```

```

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/847,185
; FILING DATE: 01-May-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/201,931
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-IM 2442
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619)535-9001
; TELEFAX: (619)535-8949
; INFORMATION FOR SEQ ID NO: 47:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 38 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 47:
US-09-847-185-47

Query Match 14.2%; Score 55.5; DB 9; Length 38;
Best Local Similarity 29.8%; Pred. No. 43;
Matches 14; Conservative 0; Mismatches 14; Indels 19; Gaps 2;

Qy 2 GFGGLGGRGKPSNEIFSRCDRCQRCFNPVVPKPLCIKICAPGCVC 48
Db 1 GCGGAGGGGCC-----CTAGC-----ACCCACCCGCTC 28

Search completed: November 17, 2003, 08:22:15
Job time : 29 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2003, 08:14:53 ; Search time 21 Seconds

(without alignments)
134.992 Million cell updates/sec

Title: US-09-506-978-1

Perfect score: 391

Sequence: 1 GFGGLGGRKCFNSNEFSR.....CRLGLENKKVCVPRSKG 67

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 215720

Minimum DB seq length: 0

Maximum DB seq length: 67

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	82	21.0	61	2	US-08-465-380-61
2	82	21.0	61	2	US-08-486-397-61
3	82	21.0	61	2	US-08-486-399-61
4	82	21.0	61	2	US-08-461-965-61
5	82	21.0	61	2	US-08-634-641-61
6	82	21.0	61	3	US-09-249-471-61
7	82	21.0	61	3	US-09-249-472-61
8	82	21.0	61	3	US-09-249-451-61
9	82	21.0	61	3	US-08-809-455-61
10	82	21.0	61	3	US-09-249-461-61
11	82	21.0	61	3	US-09-249-448-61
12	82	21.0	61	4	US-09-249-473-61
13	68	17.4	47	2	US-08-637-759B-400
14	68	17.4	47	2	US-08-871-355A-400
15	68	17.4	47	4	US-09-201-945-400
16	68	17.4	58	2	US-08-465-380-60
17	68	17.4	58	2	US-08-486-397-60
18	68	17.4	58	2	US-08-486-399-60
19	68	17.4	58	2	US-08-461-965-60
20	68	17.4	58	2	US-08-634-641-60
21	68	17.4	58	3	US-09-249-471-60
22	68	17.4	58	3	US-09-249-472-60
23	68	17.4	58	3	US-09-249-451-60
24	68	17.4	58	3	US-08-809-455-60
25	68	17.4	58	3	US-09-249-461-60
26	68	17.4	58	3	US-09-249-448-60
27	68	17.4	58	4	US-09-249-473-60

ALIGNMENTS

RESULT 1

US-08-465-380-61

; Sequence 61, Application US/08465380

; Patent No. 5863894

; GENERAL INFORMATION:

; APPLICANT: George P. Vlasuk, Patric H. Stanssens,

; APPLICANT: Joris H.L. Mensens, Marc J. Lauwereys,

; APPLICANT: Yves R. Laroche, Laurent S. Jespers,

; APPLICANT: Yannick G.J. Gansemans, Matthew Moyle,

; APPLICANT: Peter M. Bergum

; TITLE OF INVENTION: NEMATODE-EXTRACTED ANTICOAGULANT

; TITLE OF INVENTION: PROTEIN

; NUMBER OF SEQUENCES: 356

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Lyon & Lyon

; STREET: 633 West Fifth Street

; CITY: Los Angeles

; STATE: California

; COUNTRY: U.S.A.

; ZIP: 90071

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb

; MEDIUM TYPE: storage

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: IBM P.C. DOS 5.0

; SOFTWARE: Word Perfect 5.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/465,380

; FILING DATE: June 5, 1995

; CLASSIFICATION: 530

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/326,110

; FILING DATE: October 18, 1994

; ATTORNEY/AGENT INFORMATION:

; NAME: BIGGS, SUZANNE L.

; REGISTRATION NUMBER: 30,158

; REFERENCE/DOCKET NUMBER: 213/268

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (213) 489-1600

; TELEFAX: (213) 955-0440

; TELEX: 67-3510

; INFORMATION FOR SEQ ID NO: 61:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 61 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; ORIGINAL SOURCE:

; ORGANISM: Necator americanus

Sequence 5, Appli
Sequence 3, Appli
Sequence 3, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 4, Appli
Sequence 20, Appl
Sequence 47, Appl
Sequence 11, Appl
Sequence 11, Appl
Sequence 12, Appl
Sequence 1, Appl
Sequence 6, Appli
Sequence 5, Appli
Sequence 18, Appli

28 62.5 16.0 62 4 US-07-780-717C-5
29 59.5 15.2 61 2 US-08-785-530-3
30 59.5 15.2 61 2 US-09-123-850-3
31 59 15.1 61 2 US-08-785-530-1
32 59 15.1 61 2 US-09-123-850-1
33 57.5 14.7 61 2 US-08-785-530-4
34 57.5 14.7 61 2 US-09-123-850-4
35 57 14.6 10 4 US-09-394-630-20
36 55.5 14.2 38 2 US-08-902-516-47
37 55.5 14.2 38 4 US-09-847-185-47
38 55.5 14.2 45 3 US-08-899-437-11
39 55.5 14.2 45 3 US-09-126-121-11
40 55.5 14.2 46 3 US-08-915-096A-12
41 54.5 13.9 61 2 US-08-785-530-6
42 54.5 13.9 61 2 US-09-123-850-6
43 54 13.8 61 2 US-08-785-530-5
44 54 13.8 61 2 US-09-123-850-5
45 54 13.8 63 2 US-08-369-829A-18

RESULT 3

7 GENERAL INFORMATION:
7 APPLICANT: George P. Vlasuk, Patric H. Stanssens,
7 APPLICANT: Joris H.L. Mensens, Marc J. Lauwereys,
7 APPLICANT: Yves R. Laroche, Laurent S. Jespers,
7 APPLICANT: Yvanick G.J. Gansseman, Matthew Moyle,
7 APPLICANT: Peter W. Bergum
7 TITLE OF INVENTION: NEMATODE-EXTRACTED ANTICOAGULANT
7
7 TITLE OF INVENTION: PROTEIN
7
7 NUMBER OF SEQUENCES: 356

; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/249,471
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/809,455
; FILING DATE: April 17, 1997
; APPLICATION NUMBER: PCT/US95/13231
; FILING DATE: October 17, 1995
; APPLICATION NUMBER: 08/486,399
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/486,397
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/461,965
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/326,110
; FILING DATE: October 18, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BIGGS, SUZANNE L.
; REGISTRATION NUMBER: 30,158
; REFERENCE/DOCKET NUMBER: 216/270
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 61:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 61 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Necator americanus
; US-09-249-471-61

Query Match 21.0%; Score 82; DB 3; Length 61;
Best Local Similarity 30.9%; Pred. No. 0.066;
Matches 17; Conservative 9; Mismatches 19; Indels 10; Gaps 2;

QY 12 CPSNEIFSRCDGRCQRCFPCNVVVKPLCIKICAPGCV-----CRLGYLRNKKKVCV 61
Db 4 CPANEEWRECGTPEPKCNQMP-----DICTMNCIVDVQCCKEGYKRHETKGCL 53

RESULT 7

US-09-249-472-61
; Sequence 61, Application US/09249472
; Patent No. 6046318
; GENERAL INFORMATION:
; APPLICANT: Vlasuk, George Phillip
; APPLICANT: Stanssens, Patrick Eric Hugo
; APPLICANT: Messens, Joris Hilda Lieven
; APPLICANT: Lauwereys, Marc Josef
; APPLICANT: Laroche, Yves Rene
; APPLICANT: Jespers, Laurent Stephane
; APPLICANT: Ganssemans, Yannick Georges Jozef
; APPLICANT: Moyle, Matthew
; APPLICANT: Bergum, Peter W.
; TITLE OF INVENTION: NEMATODE-EXTRACTED SERINE PROTEASE
; TITLE OF INVENTION: INHIBITORS AND ANTICOAGULANT
; NUMBER OF SEQUENCES: 356
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon

; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/249,472
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/809,455
; FILING DATE: April 17, 1997
; APPLICATION NUMBER: PCT/US95/13231
; FILING DATE: October 17, 1995
; APPLICATION NUMBER: 08/486,399
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/486,397
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/465,380
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/461,965
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/326,110
; FILING DATE: October 18, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BIGGS, SUZANNE L.
; REGISTRATION NUMBER: 30,158
; REFERENCE/DOCKET NUMBER: 216/270
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 61:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 61 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Necator americanus
; US-09-249-472-61

Query Match 21.0%; Score 82; DB 3; Length 61;
Best Local Similarity 30.9%; Pred. No. 0.066;
Matches 17; Conservative 9; Mismatches 19; Indels 10; Gaps 2;

QY 12 CPSNEIFSRCDGRCQRCFPCNVVVKPLCIKICAPGCV-----CRLGYLRNKKKVCV 61
Db 4 CPANEEWRECGTPEPKCNQMP-----DICTMNCIVDVQCCKEGYKRHETKGCL 53

RESULT 8

US-09-249-451-61
; Sequence 61, Application US/09249451
; Patent No. 6087487
; GENERAL INFORMATION:
; APPLICANT: Vlasuk, George Phillip
; APPLICANT: Stanssens, Patrick Eric Hugo
; APPLICANT: Messens, Joris Hilda Lieven
; APPLICANT: Lauwereys, Marc Josef
; APPLICANT: Laroche, Yves Rene
; APPLICANT: Jespers, Laurent Stephane
; APPLICANT: Ganssemans, Yannick Georges Jozef
; APPLICANT: Moyle, Matthew
; APPLICANT: Bergum, Peter W.
; TITLE OF INVENTION: NEMATODE-EXTRACTED SERINE PROTEASE
; TITLE OF INVENTION: INHIBITORS AND ANTICOAGULANT

;
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 356
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/249,451
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/809,455
; FILING DATE: April 17, 1997
; APPLICATION NUMBER: PCT/US95/13231
; FILING DATE: October 17, 1995
; APPLICATION NUMBER: 08/486,399
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/486,397
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/465,380
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/461,965
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/326,110
; FILING DATE: October 18, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BIGGS, SUZANNE L.
; REGISTRATION NUMBER: 30,158
; REFERENCE/DOCKET NUMBER: 216/270
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 61:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 61 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Necator americanus
; US-09-249-451-61

Query Match 21.0%; Score 82; DB 3; Length 61;
Best Local Similarity 30.9%; Pred.No. 0.066;
Matches 17; Conservative 9; Mismatches 19; Indels 10; Gaps 2;

QY 12 CPSEIFSRCDGRCQFCPNVVKPLCIKICAPGCV-----CRLGYLRNKKYCV 61
Db 4 CPANEEWREGCTPECPKCNQMP-----DICTMNCIVDVQCKEGYKRHETKGL 53

RESULT 9
US-08-809-455-61
; Sequence 61, Application US/08809455
; Patent No. 6090916
; GENERAL INFORMATION:
; APPLICANT: Vlasuk, George Phillip
; APPLICANT: Stanssens, Patrick Eric Hugo
; APPLICANT: Messens, Joris Hilda Lieven
; APPLICANT: Lauwereys, Marc Josef
; APPLICANT: Laroche, Yves Rene
; APPLICANT: Jespers, Laurent Stephane
; APPLICANT: Gansemans, Yannick Georges Jozef

;
; APPLICANT: Moyle, Matthew
; APPLICANT: Beigum, Peter W.
; TITLE OF INVENTION: NEMATODE-EXTRACTED SERINE PROTEASE
; TITLE OF INVENTION: INHIBITORS AND ANTICOAGULANT
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 356
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/809,455
; FILING DATE: April 17, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/13231
; FILING DATE: October 17, 1995
; APPLICATION NUMBER: 08/486,399
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/486,397
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/465,380
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/461,965
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/326,110
; FILING DATE: October 18, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BIGGS, SUZANNE L.
; REGISTRATION NUMBER: 30,158
; REFERENCE/DOCKET NUMBER: 216/270
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 61:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 61 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Necator americanus
; US-08-809-455-61

Query Match 21.0%; Score 82; DB 3; Length 61;
Best Local Similarity 30.9%; Pred.No. 0.066;
Matches 17; Conservative 9; Mismatches 19; Indels 10; Gaps 2;

QY 12 CPSEIFSRCDGRCQFCPNVVKPLCIKICAPGCV-----CRLGYLRNKKYCV 61
Db 4 CPANEEWREGCTPECPKCNQMP-----DICTMNCIVDVQCKEGYKRHETKGL 53

RESULT 10
US-09-249-461-61
; Sequence 61, Application US/09249461
; Patent No. 6096877
; GENERAL INFORMATION:
; APPLICANT: Vlasuk, George Phillip
; APPLICANT: Stanssens, Patrick Eric Hugo
; APPLICANT: Messens, Joris Hilda Lieven
; APPLICANT: Lauwereys, Marc Josef
; APPLICANT: Laroche, Yves Rene

Query Match 21.0%; Score 82; DB 4; Length 61;
Best Local Similarity 30.9%; Pred. No. 0.066;
Matches 17; Conservative 9; Mismatches 19; Indels 10; Gaps 2;

Query Match 17.4%; Score 68; DB 2; Length 47;
Best Local Similarity 55.6%; Pred. No. 1.4;
Matches 10; Conservative 2; Mismatches 6; Indels

```

RESULT 14
US-08-871-355A-400
; Sequence 400, Application US/08871355A
; Patent No. 6015669
; GENERAL INFORMATION:
; APPLICANT: David William Holden
; TITLE OF INVENTION: Identification of Genes
; NUMBER OF SEQUENCES: 501
; CORRESPONDENCE ADDRESS:
; ADDRESS: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; STREET: 1201 West Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309-3450

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/871,355A
FILING DATE: 09-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB95/02875
FILING DATE: 11-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: RPMS 101 CON
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404) 873-8794
TELEFAX: (404) 873-8795
INFORMATION FOR SEQ ID NO: 400:
SEQUENCE CHARACTERISTICS:
LENGTH: 47 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-08-871-355A-400

Query Match 17.4%; Score 68; DB 3; Length 47;
Best Local Similarity 55.6%; Pred. No. 1.4;
Matches 10; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 12 CPSNEIFSRCDGRCQRCFC 29
|||: : |||||
Db 30 CPAGKPLSRCDGRCDEIC 47

RESULT 15
US-09-201-945-400
Sequence 400, Application US/09201945
Patent No. 6342215
GENERAL INFORMATION:
APPLICANT: David William Holden
TITLE OF INVENTION: Identification of Genes
NUMBER OF SEQUENCES: 501
CORRESPONDENCE ADDRESS:
ADDRESSER: Patrea L. Pabst
STREET: 2800 One Atlantic Center
STREET: 1201 West Peachtree Street
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/201,945
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/637,759
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: RPMS 101
TELECOMMUNICATION INFORMATION:

TELEPHONE: (404) 873-8794
TELEFAX: (404) 873-8795
INFORMATION FOR SEQ ID NO: 400:
SEQUENCE CHARACTERISTICS:
LENGTH: 47 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-09-201-945-400

Query Match 17.4%; Score 68; DB 4; Length 47;
Best Local Similarity 55.6%; Pred. No. 1.4;
Matches 10; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 12 CPSNEIFSRCDGRCQRCFC 29
|||: : |||||
Db 30 CPAGKPLSRCDGRCDEIC 47

Search completed: November 17, 2003, 08:17:59
Job time : 22 secs